

Converse Bounds for Entropy-Constrained Quantization Via a Variational Entropy Inequality

Tobias Koch and Gonzalo Vazquez-Vilar

Abstract

We derive a lower bound on the smallest output entropy that can be achieved via vector quantization of a d -dimensional source with given expected r th-power distortion. Specialized to the one-dimensional case, and in the limit of vanishing distortion, this lower bound converges to the output entropy achieved by a uniform quantizer, thereby recovering the result by Gish and Pierce that uniform quantizers are asymptotically optimal as the allowed distortion tends to zero. Our lower bound holds for all d -dimensional memoryless sources having finite differential entropy and whose integer part has finite entropy. In contrast to Gish and Pierce, we do not require any additional constraints on the continuity or decay of the source probability density function. For one-dimensional sources, the derivation of the lower bound reveals a necessary condition for a sequence of quantizers to be asymptotically optimal as the allowed distortion tends to zero. This condition implies that any sequence of asymptotically-optimal almost-regular quantizers must converge to a uniform quantizer as the allowed distortion tends to zero.

I. INTRODUCTION

Suppose we wish to quantize a memoryless source with an r th-power distortion not larger than D . More specifically, suppose a source produces the sequence of independent and identically distributed, d -dimensional, real-valued vectors $\{\mathbf{X}_k, k \in \mathbb{Z}\}$ according to the distribution $P_{\mathbf{X}}$ and we employ a vector quantizer that produces a sequence of quantized symbols $\{\hat{\mathbf{X}}_k, k \in \mathbb{Z}\}$ satisfying

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\|\mathbf{X}_k - \hat{\mathbf{X}}_k\|^r \right] \leq D \quad (1)$$

for some norm $\|\cdot\|$ and some exponent $r > 0$. (We use $\overline{\lim}$ to denote the *limit superior* and $\underline{\lim}$ to denote the *limit inferior*.) Rate-distortion theory states that if for every blocklength n and distortion constraint D we quantize the sequence of source vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$ to one of e^{nR} possible sequences of quantized symbols $\hat{\mathbf{X}}_1, \dots, \hat{\mathbf{X}}_n$, then

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 714161), from the 7th European Union Framework Programme under Grant 333680, from the Ministerio de Economía y Competitividad of Spain under Grants TEC2013-41718-R, RYC-2014-16332, IJCI-2015-27020, TEC2015-69648-REDC, and TEC2016-78434-C3-3-R (AEI/FEDER, EU), and from the Comunidad de Madrid under Grant S2103/ICE-2845. The material in this paper was presented in part at the 2016 IEEE International Symposium on Information Theory, Barcelona, Spain, July 2016.

The authors are with the Signal Theory and Communications Department, Universidad Carlos III de Madrid, 28911, Leganés, Spain and also with the Gregorio Marañón Health Research Institute (e-mails: koch@tsc.uc3m.es and gvazquez@tsc.uc3m.es).

the smallest rate R (in nats per source symbol) for which there exists a vector quantizer satisfying (1) is given by [1]

$$R(D) = \inf_{P_{\hat{\mathbf{X}}|\mathbf{X}}} I(\mathbf{X}; \hat{\mathbf{X}}) \quad (2)$$

where the infimum is over all conditional distributions of $\hat{\mathbf{X}}$ given \mathbf{X} for which

$$\mathbb{E} [\|\mathbf{X} - \hat{\mathbf{X}}\|^r] \leq D \quad (3)$$

and where the expectation in (3) is computed with respect to the joint distribution $P_{\mathbf{X}}P_{\hat{\mathbf{X}}|\mathbf{X}}$. Here and throughout this paper we omit the time indices where they are immaterial. The rate $R(D)$ as a function of D is referred to as the *rate-distortion function*.

While $R(D)$ characterizes the rate of the best vector quantizer that quantizes the source with r th-power distortion not exceeding D , sometimes quantizing blocks of n source symbols may not be feasible, especially if n is large (which is typically required to achieve (2)). In this case, it might be more practical to quantize each source symbol separately using a vector quantizer, defined as a (deterministic) mapping $q(\cdot)$ from the source alphabet \mathcal{X} to the (countable) reconstruction alphabet $\hat{\mathcal{X}}$.

In this paper, we consider the symbol-wise quantization of d -dimensional source vectors. This setup is sufficiently general to comprise various problems of interest in high-resolution vector quantization. For example, it allows us to analyze the performance of quantization schemes that buffer d consecutive symbols of a one-dimensional memoryless source and then quantize them using a d -dimensional vector quantizer. Furthermore, the quantization of stationary sources with memory can be studied by combining the analysis of symbol-wise, d -dimensional quantization with a limiting argument where $d \rightarrow \infty$.

We define the rate of the vector quantizer as the entropy of the quantized source symbol $\hat{\mathbf{X}} = q(\mathbf{X})$. Thus, the smallest rate of a symbol-wise quantizer satisfying the distortion constraint D is given by

$$R_{r,d}(D) \triangleq \inf_{q(\cdot)} H(q(\mathbf{X})) \quad (4)$$

where the infimum is over the set of quantizers $q(\cdot)$ satisfying (3). Since \mathbf{X} determines the quantizer output $q(\mathbf{X})$, we have $H(q(\mathbf{X})|\mathbf{X}) = 0$ and the rate $R_{r,d}(D)$ can be written in the same form as (2) but with $P_{\hat{\mathbf{X}}|\mathbf{X}}$ replaced by $q(\cdot)$:

$$R_{r,d}(D) = \inf_{q(\cdot)} I(\mathbf{X}; q(\mathbf{X})). \quad (5)$$

Since $\hat{\mathbf{X}} = q(\mathbf{X})$ corresponds to a deterministic $P_{\hat{\mathbf{X}}|\mathbf{X}}$, it follows that $R_{r,s}(D) \geq R(D)$.

Any discrete memoryless source can be losslessly described by a variable-length code whose expected length is roughly the entropy of the source [2], [3]. Consequently, $R_{r,d}(D)$ is the smallest expected length of a vector quantization scheme that first quantizes each source symbol using a vector quantizer and then compresses the resulting sequence of quantized symbols using a lossless variable-length code.

In this paper, we focus on the asymptotic rate-distortion tradeoff in the limit as the permitted distortion tends to zero. Specifically, we study the asymptotic excess rate with respect to the rate-distortion function defined as

$$\mathbf{R}_{r,d} \triangleq \lim_{D \downarrow 0} \{R_{r,d}(D) - R(D)\}. \quad (6)$$

For one-dimensional sources ($d = 1$) and quadratic distortion ($r = 2$), Gish and Pierce demonstrated that the excess rate is equal to [4]

$$\mathbb{R}_{2,1} = \frac{1}{2} \log \frac{\pi e}{6} \quad (7)$$

where $\log(\cdot)$ denotes the natural logarithm. They further showed that this excess rate can be achieved by a uniform quantizer, hence the well-known result that “uniform quantizers are asymptotically optimal as the allowed distortion tends to zero.”¹ For multi-dimensional sources, only bounds on $\mathbb{R}_{r,d}$ are available. To obtain (7), Gish and Pierce [4] imposed constraints on the continuity and decay of the probability density function (pdf) of \mathbf{X} . Furthermore, they merely provide an intuitive explanation of their converse result together with an outline of the proof—at the end of [4, Appendix II] they write “The complete proof is surprisingly long and will not be given here.”

The result (7) is equivalent to a result by Zador [6], which concerns the asymptotic excess distortion with respect to the distortion-rate function as the rate tends to infinity. Indeed, let $D_{r,d}(R)$ denote the minimum distortion achievable with a symbol-wise quantizer whose output has an entropy not exceeding R , i.e.,

$$D_{r,d}(R) \triangleq \inf_{q(\cdot)} \mathbf{E} [\|\mathbf{X} - q(\mathbf{X})\|^r] \quad (8)$$

where the infimum is over the set of quantizers $q(\cdot)$ satisfying $H(q(\mathbf{X})) \leq R$. Zador’s theorem states that

$$\lim_{R \rightarrow \infty} e^{\frac{r}{d}R} D_{r,d}(R) = b_{r,d} e^{\frac{r}{d}h(\mathbf{X})} \quad (9)$$

where $b_{r,d}$ is a constant that only depends on r and d but not on the distribution of \mathbf{X} . Zador did not evaluate the constant $b_{r,d}$, but he did provide upper and lower bounds on $b_{r,d}$ that become tight for large d . Furthermore, for one-dimensional sources and quadratic distortion, it can be shown that $b_{2,1} = 1/12$. Taking logarithms on both sides of (9), and replacing $R \leftrightarrow R_{r,d}(D)$ and $D_{r,d}(R) \leftrightarrow D$, we thus obtain that

$$R_{2,1}(D) = h(X) + \frac{1}{2} \log \frac{1}{D} - \frac{1}{2} \log 12 + o_R(1) \quad (10)$$

where $o_R(1)$ denotes error terms that vanish as R tends to infinity. Furthermore, the rate-distortion function can be approximated as [7]–[9]

$$R(D) = h(X) + \frac{1}{2} \log \frac{1}{D} - \frac{1}{2} \log(2\pi e) + o_D(1) \quad (11)$$

where $o_D(1)$ denotes error terms that vanish as D tends to zero. Hence, the equivalence of Zador’s theorem (9) and Gish and Pierce’s result (7) follows by applying (10) and (11) to (6).

While Zador’s original proof of (9) was flawed, a rigorous proof for quadratic distortion was given by Gray, Linder, and Li by using a Lagrangian formulation of variable-rate vector quantization [10]. Their proof follows Zador’s approach of 1) proving the result for sources with a uniform pdf on the unit cube; 2) extending it to piecewise constant pdfs on disjoint cubes of equal sides; 3) proving the result for a general pdf on a cube; and 4) proving the result for general pdfs. Gray *et al.* do not impose any constraints on the continuity or decay of the pdf of \mathbf{X} , so their proof is more general than the proofs by Zador [6] and by Gish and Pierce [4].

¹The fact that, in the high-resolution case, the expected quadratic distortion of uniform scalar quantization exceeds the least distortion achievable by any quantization scheme by a factor of only $\pi e/6$ was already discovered by Koshelev in 1963. See [5] and references therein for more details.

In this paper, we derive a lower bound on $R_{r,d}$ that recovers (7) for one-dimensional sources and quadratic distortion. In contrast to [10], our proof follows essentially along the lines outlined by Gish and Pierce [4]. We do not impose any constraints on the continuity or decay of the pdf of \mathbf{X} , so our proof is as general as the proof by Gray *et al.*, and it is more general than the proof by Gish and Pierce.

For one-dimensional sources, the derivation of the lower bound reveals a necessary condition for a sequence of quantizers (parametrized by D) to achieve the asymptotic excess rate $R_{r,1}$. We apply this condition to the family of *almost-regular quantizers*, which was introduced by György and Linder in [11] and includes the uniform quantizers. Almost-regular quantizers are relevant because they achieve $D_{r,1}(R)$ when $r \geq 1$ [11, Theorem 3]. Thus, for one-dimensional sources and r th-power distortion with $r \geq 1$, we can restrict ourselves to almost-regular quantizers without loss of optimality. The necessary condition implies that any sequence of almost-regular quantizers achieving $R_{r,1}$ must converge to a uniform quantizer as $D \rightarrow 0$. This suggests that asymptotically-optimal quantizers must essentially be uniform.

The rest of this paper is organized as follows. Section II introduces the problem setup and presents the main result of this paper, Theorem 1. Section III provides a back-of-the-envelope derivation of Theorem 1 that serves as an outline for the proof. Section IV contains the complete proof of this theorem. Section V presents a necessary condition for a sequence of quantizers to achieve the asymptotic excess rate. Section VI assesses the tightness of the lower bound presented in Theorem 1 for multi-dimensional sources by numerically comparing it to several upper bounds achievable by lattice quantizers. Section VII concludes the paper with a summary and discussion of the results.

II. PROBLEM SETUP AND MAIN RESULT

We consider a d -dimensional, real-valued source \mathbf{X} with support $\mathcal{X} \subseteq \mathbb{R}^d$ whose distribution is absolutely continuous with respect to the Lebesgue measure, and we denote its pdf by $f_{\mathbf{X}}$. We require the source to satisfy the following two conditions:

C1 $\mathbf{x} \mapsto f_{\mathbf{X}}(\mathbf{x}) \log f_{\mathbf{X}}(\mathbf{x})$ is integrable, ensuring that the differential entropy

$$h(\mathbf{X}) \triangleq - \int_{\mathcal{X}} f_{\mathbf{X}}(\mathbf{x}) \log f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \quad (12)$$

is well-defined and finite;

C2 the integer part of the source \mathbf{X} has finite entropy, i.e.,

$$H(\lfloor \mathbf{X} \rfloor) < \infty. \quad (13)$$

Here $\lfloor \mathbf{a} \rfloor$, $\mathbf{a} = (a_1, \dots, a_d) \in \mathbb{R}^d$ denotes the element-wise floor function, i.e., $\lfloor \mathbf{a} \rfloor = (\lfloor a_1 \rfloor, \dots, \lfloor a_d \rfloor)$ where $\lfloor a_\ell \rfloor$, $\ell = 1, \dots, d$ denotes the largest integer not larger than a_ℓ .

Condition C2 requires that quantizing the source with a cubic lattice quantizer of unit-volume cells gives rise to a discrete random variable of finite entropy. This is necessary for the asymptotic excess rate $R_{r,d}$ to be well-defined. Indeed, as demonstrated in [9], if $H(\lfloor \mathbf{X} \rfloor) = \infty$ then the rate-distortion function $R(D)$ is infinite for any finite D . Since $R_{r,d}(D) \geq R(D)$, this implies that in this case $R_{r,s}(D) - R(D)$ is of the form $\infty - \infty$. Fortunately,

Condition C2 is very mild. For example, by generalizing [12, Proposition 1] to the vector case, it can be shown that it is satisfied if $\mathbb{E}[\log(1 + \|\mathbf{X}\|)] < \infty$. This in turn is true, for example, for sources for which $\mathbb{E}[\|\mathbf{X}\|^\alpha] < \infty$ for some $\alpha > 0$.

The quantity $H(\lfloor \mathbf{X} \rfloor)$ is intimately related with the *Rényi information dimension* defined in [13]; see also [12], [14]. Indeed, it can be shown that a source vector has finite Rényi information dimension if, and only if, (13) is satisfied [12, Proposition 1].

The quantizer is characterized by the (Borel measurable) function $q: \mathcal{X} \rightarrow \hat{\mathcal{X}}$ for some countable reconstruction alphabet $\hat{\mathcal{X}} \subseteq \mathbb{R}^d$. Equivalently, we characterize $q(\cdot)$ by the quantization regions \mathcal{S}_i , $i \in \mathbb{Z}$ and corresponding reconstruction values $\hat{\mathbf{x}}_i$, $i \in \mathbb{Z}$. Specifically, \mathcal{S}_i , $i \in \mathbb{Z}$ are disjoint (Borel measurable) subsets of \mathbb{R}^d that together with the reconstruction values $\hat{\mathbf{x}}_i$, $i \in \mathbb{Z}$ satisfy

$$\bigcup_i \mathcal{S}_i = \mathcal{X} \quad (14a)$$

$$q(\mathbf{x}) = \sum_i \hat{\mathbf{x}}_i \mathbb{1}\{\mathbf{x} \in \mathcal{S}_i\}, \quad \text{for } \mathbf{x} \in \mathcal{X} \quad (14b)$$

where $\mathbb{1}\{\cdot\}$ denotes the indicator function. To simplify notation, we denote the Lebesgue measure of the quantization region \mathcal{S}_i by Δ_i and the probability of \mathbf{X} being in \mathcal{S}_i by p_i .

The main result of this paper is a lower bound on the excess rate $R_{r,d}$ for general r and d . For one-dimensional sources and quadratic distortion, it recovers the excess rate (7) by Gish and Pierce. However, in contrast to Gish and Pierce's result, our bound does not require any continuity or decay conditions on the behavior of the source pdf—it holds for all source vectors having a pdf, having finite differential entropy, and having finite Rényi information dimension.

Theorem 1 (Main Result): Let the source vector \mathbf{X} have a pdf, and assume that $h(\mathbf{X})$ and $H(\lfloor \mathbf{X} \rfloor)$ are finite. Then, the excess rate $R_{r,d}$, as defined in (6), is lower-bounded by

$$R_{r,d} \geq \frac{d}{r} \log \left(\frac{\Gamma(1 + d/r)^{r/d} e}{1 + d/r} \right) \quad (15)$$

where $\Gamma(\cdot)$ denotes the Gamma function.

Proof: See Section IV. ■

In the one-dimensional case, (15) becomes

$$R_{r,1} \geq \frac{1}{r} \log \left(\frac{\Gamma(1 + 1/r)^r e}{1 + 1/r} \right). \quad (16)$$

As we shall see next, (16) can be achieved by a uniform quantizer, so in the one-dimensional case the lower bound (15) is tight. Furthermore, for quadratic distortion, (16) is equal to $1/2 \log(\pi e/6)$, hence it recovers the excess rate obtained by Gish and Pierce.

To demonstrate the tightness of (15) in the one-dimensional case, and to assess the accuracy of (15) in higher-dimensional cases, we consider an upper bound on the excess rate that follows by restricting ourselves to the class of *tessellating quantizers*. A polytope \mathcal{P} is *tessellating* if there exists a partition of \mathbb{R}^d consisting of translated and/or rotated copies of \mathcal{P} ; a *tessellating quantizer*, denoted by $q_{\mathcal{P}}: \mathcal{X} \rightarrow \hat{\mathcal{X}}$, is a quantizer whose quantization regions \mathcal{S}_i are translated and/or rotated copies of a tessellating convex polytope \mathcal{P} and the corresponding reconstruction

values $\hat{\mathbf{x}}_i$ are the centroids of \mathcal{S}_i . A special case of a tessellating quantizer is a *lattice quantizer*, i.e., a quantizer whose quantization regions are the Voronoi cells of a d -dimensional lattice. Note that in the one-dimensional case the only convex polytope is the interval, so in this case the tessellating quantizer is the uniform quantizer. For the class of tessellating quantizers, Linder and Zeger [15] derived an asymptotic expression equivalent to (9).

Theorem 2 (Linder and Zeger [15, Theorem 1]): Let the source vector \mathbf{X} have a pdf, and assume that $h(\mathbf{X})$ and $H(\lfloor \mathbf{X} \rfloor)$ are finite. Then, a tessellating quantizer $q_{\mathcal{P}}(\cdot)$ with r th-power distortion $\mathbb{E}[\|\mathbf{X} - q_{\mathcal{P}}(\mathbf{X})\|^r] = D$ and rate $R_{\mathcal{P}}(D) \triangleq H(q_{\mathcal{P}}(\mathbf{X}))$ satisfies

$$\lim_{D \downarrow 0} D e^{\frac{r}{d} R_{\mathcal{P}}(D)} = \ell(\mathcal{P}) e^{\frac{r}{d} h(\mathbf{X})} \quad (17)$$

where $\ell(\mathcal{P})$ denotes the normalized r -th moment of \mathcal{P} , defined as

$$\ell(\mathcal{P}) \triangleq \frac{\int_{\mathcal{P}} \|\mathbf{x} - \hat{\mathbf{x}}\|^r d\mathbf{x}}{V(\mathcal{P})^{1+r/d}} \quad (18)$$

and $V(\mathcal{P})$ denotes the volume of \mathcal{P} .

Remark: To be precise, [15, Theorem 1] requires that $H(q_{\mathcal{P}_{\alpha}}(\mathbf{X})) < \infty$ for some $\alpha > 0$ rather than (13), i.e., $H(\lfloor \mathbf{X} \rfloor) < \infty$. (Here, $\mathcal{P}_{\alpha} = \{x \in \mathbb{R}^d : x/\alpha \in \mathcal{P}\}$ denotes the polytope \mathcal{P} rescaled by α .) Nevertheless, its proof hinges on a lemma by Csiszár (cf. [15, Lemma 2]), which also applies if the condition $H(q_{\mathcal{P}_{\alpha}}(\mathbf{X})) < \infty$ is replaced by (13). Specifically, by setting in [15, Lemma 2] the partition $\mathcal{B}_0 = \{B_1, B_2, \dots\}$ of \mathbb{R}^d to be the set of d -dimensional cubes of unit-volume with the lower-most cornerpoint located at coordinates $\mathbf{i} \in \mathbb{Z}^d$, this partition satisfies the lemma's conditions provided that (13) holds.

Taking logarithms on both sides of (17), we obtain

$$R_{\mathcal{P}}(D) = h(\mathbf{X}) + \frac{d}{r} \log \frac{1}{D} + \frac{d}{r} \log \ell(\mathcal{P}) + o_D(1). \quad (19)$$

Since a tessellating quantizer with r th-power distortion D satisfies (3), the rate $R_{\mathcal{P}}(D)$ upper-bounds $R_{r,d}(D)$. Furthermore, the rate-distortion function $R(D)$ can be lower-bounded as [16]

$$R(D) \geq h(\mathbf{X}) + \frac{d}{r} \log \frac{1}{D} - \frac{d}{r} \log \left(\frac{r}{d} (V_d \Gamma(1 + d/r))^{r/d} e \right) \quad (20)$$

where V_d denotes the volume of the unit ball $\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| \leq 1\}$. The right-hand side (RHS) of (20) is referred to as *Shannon lower bound*. It has been demonstrated that its difference to $R(D)$ vanishes as D tends to zero, provided that the source distribution satisfies certain conditions; see, e.g., [7]–[9]. A finite-blocklength refinement of this bound can be found in [17], [18]. Recently, it has been demonstrated that for sources with finite differential entropy the Shannon lower bound is asymptotically tight if, and only if, $H(\lfloor \mathbf{X} \rfloor)$ is finite [9]. Thus, we have

$$R(D) = h(\mathbf{X}) + \frac{d}{r} \log \frac{1}{D} - \frac{d}{r} \log \left(\frac{r}{d} (V_d \Gamma(1 + d/r))^{r/d} e \right) + o_D(1) \quad (21)$$

for the class of sources considered in this paper.

Combining (19) with (21), we obtain

$$\lim_{D \downarrow 0} \{R_{\mathcal{P}}(D) - R(D)\} = \frac{d}{r} \log \left(\frac{r}{d} (V_d \Gamma(1 + d/r))^{r/d} e \right) + \frac{d}{r} \log \ell(\mathcal{P}). \quad (22)$$

Recalling that $R_{r,d}(D) \leq R_{\mathcal{P}}(D)$ for every D , this yields

$$R_{r,d} \leq \frac{d}{r} \log \left(\frac{r}{d} (V_d \Gamma(1 + d/r))^{r/d} e \right) + \inf_{\mathcal{P}} \frac{d}{r} \log \ell(\mathcal{P}) \quad (23)$$

where the infimum is over all d -dimensional, tessellating, convex polytopes \mathcal{P} .

Using that in the one-dimensional case the only convex polytope is the interval, and noting that the interval has the normalized r -th moment

$$\ell(\mathcal{P}) = \frac{1}{2^r(1+r)} \quad (24)$$

the upper bound (23) becomes in this case

$$\mathbb{R}_{r,1} \leq \frac{1}{r} \log \left(\frac{\Gamma(1+1/r)^r e}{1+1/r} \right) \quad (25)$$

which coincides with (16). Thus, in the one-dimensional case a tessellating quantizer (which in this case is the uniform quantizer) is asymptotically optimal.

III. DERIVATION FOR ONE-DIMENSIONAL SOURCES AND CERTAIN QUANTIZERS

Before proving Theorem 1, we provide a simplified derivation of the lower bound (15) for one-dimensional sources ($d = 1$) and quadratic distortion ($r = 2$) that will serve as an outline for the complete proof of Theorem 1 given in Section IV. Particularized to this setting, Theorem 1 becomes

$$\mathbb{R}_{2,1} \geq \frac{1}{2} \log \frac{\pi e}{6}. \quad (26)$$

In our derivation we shall only consider quantizers satisfying

$$\sup_i \sup_{x \in \mathcal{S}_i} (x - \hat{x}_i)^2 \leq \alpha D, \quad \text{for some constant } \alpha. \quad (27)$$

This simplifying assumption is, for example, satisfied by the uniform quantizer when \hat{x}_i is the midpoint of \mathcal{S}_i and the cell length Δ vanishes proportionally to \sqrt{D} . However, it is *prima facie* unclear whether (27) holds without loss of optimality for general sources.

By (5), we have

$$R_{2,1}(D) = \inf_{q(\cdot)} I(X; \hat{X}) = h(X) - \sup_{q(\cdot)} h(X|\hat{X}). \quad (28)$$

We upper-bound $h(X|\hat{X})$ by using that, conditioned on $\hat{X} = \hat{x}_i$, the support of X is \mathcal{S}_i , so a uniform distribution over \mathcal{S}_i maximizes the differential entropy [2, Theorem 11.1.1]:

$$h(X|\hat{X} = \hat{x}_i) \leq \log \Delta_i. \quad (29)$$

Averaging over \hat{X} then yields

$$R_{2,1}(D) \geq h(X) - \sup_{q(\cdot)} \sum_i p_i \log \Delta_i. \quad (30)$$

By Jensen's inequality, this can be further lower-bounded by

$$R_{2,1}(D) \geq h(X) - \frac{1}{2} \log \left(\sup_{q(\cdot)} \sum_i p_i \Delta_i^2 \right). \quad (31)$$

Together with (11), this yields

$$\lim_{D \downarrow 0} \{R_{2,1}(D) - R(D)\} \geq \lim_{D \downarrow 0} \left\{ \frac{1}{2} \log D + \frac{1}{2} \log(2\pi e) - \frac{1}{2} \log \left(\sup_{q(\cdot)} \sum_i p_i \Delta_i^2 \right) \right\}. \quad (32)$$

In order to prove (26), it remains to show that, for any sequence of quantizers (parametrized by D),

$$\overline{\lim}_{D \downarrow 0} \frac{1}{D} \sum_i p_i \Delta_i^2 \leq 12. \quad (33)$$

Then the RHS of (32) is lower-bounded by $1/2 \log(\pi e/6)$ and we obtain (26) upon noting that the left-hand side (LHS) of (32) is equal to $R_{2,1}$. Hence we recover Theorem 1 for one-dimensional sources and quadratic distortion.

The upper bound (33) follows along the lines of the proof of [15, Lemma 1]. We first express $\mathbb{E}[(X - \hat{X})^2]$ as

$$\begin{aligned} \mathbb{E}[(X - \hat{X})^2] &= \sum_i \int_{\mathcal{S}_i} f_X(x) (x - \hat{x}_i)^2 dx \\ &= \sum_i p_i \frac{1}{\Delta_i} \int_{\mathcal{S}_i} (x - \hat{x}_i)^2 dx - \sum_i \int_{\mathcal{S}_i} \left[\frac{p_i}{\Delta_i} - f_X(x) \right] (x - \hat{x}_i)^2 dx. \end{aligned} \quad (34)$$

We next note that the region \mathcal{S}_i of measure Δ_i that minimizes $\int_{\mathcal{S}_i} (x - \hat{x}_i)^2 dx$ is the interval $[\hat{x}_i - \frac{\Delta_i}{2}, \hat{x}_i + \frac{\Delta_i}{2}]$, so

$$\frac{1}{\Delta_i} \int_{\mathcal{S}_i} (x - \hat{x}_i)^2 dx \geq \frac{\Delta_i^2}{12}. \quad (35)$$

The first term on the RHS of (34) can therefore be lower-bounded by

$$\sum_i p_i \frac{1}{\Delta_i} \int_{\mathcal{S}_i} (x - \hat{x}_i)^2 dx \geq \sum_i p_i \frac{\Delta_i^2}{12}. \quad (36)$$

To evaluate the second term on the RHS of (34), we introduce the piecewise-constant pdf

$$f_X^{(\Delta)}(x) \triangleq \sum_i \frac{p_i}{\Delta_i} \mathbb{1}_{\{x \in \mathcal{S}_i\}}, \quad x \in \mathbb{R}. \quad (37)$$

With this, we can upper-bound the second term on the RHS of (34) as

$$\begin{aligned} \sum_i \int_{\mathcal{S}_i} \left[\frac{p_i}{\Delta_i} - f_X(x) \right] (x - \hat{x}_i)^2 dx &= \sum_i \int_{\mathcal{S}_i} \left[f_X^{(\Delta)}(x) - f_X(x) \right] (x - \hat{x}_i)^2 dx \\ &\leq \alpha D \int \left| f_X^{(\Delta)}(x) - f_X(x) \right| dx \end{aligned} \quad (38)$$

since, by (27), we have $\sup_i \sup_{x \in \mathcal{S}_i} (x - \hat{x}_i)^2 \leq \alpha D$.

By Lebesgue's differentiation theorem, $f_X^{(\Delta)}$ converges to f_X almost everywhere as $\sup_i \Delta_i \rightarrow 0$. It therefore follows from Scheffe's Lemma [19, Theorem 16.12] that

$$\lim_{D \downarrow 0} \int \left| f_X^{(\Delta)}(x) - f_X(x) \right| dx = 0. \quad (39)$$

Combining (36) and (38) with (34), and using that $\mathbb{E}[(X - \hat{X})^2] \leq D$, we obtain

$$\sum_i p_i \Delta_i^2 \leq 12D \left(1 + \alpha \int \left| f_X^{(\Delta)}(x) - f_X(x) \right| dx \right). \quad (40)$$

Together with (39) this proves (33).

IV. PROOF OF THEOREM 1

A. Variational Entropy Inequality and Auxiliary Results

The above back-of-the-envelope derivation directly generalizes to multi-dimensional sources and r th-power distortion. In order to prove Theorem 1, it would remain to show that (27) holds without loss of optimality. Unfortunately, for general sources this appears to be a difficult task. Indeed, the quantization regions of the optimal quantizer are difficult to characterize since the optimal quantizer (and hence the number of quantization regions together with their locations and volumes) changes with D . To sidestep this problem, we replace (29) by an upper bound on $h(\mathbf{X}|\hat{\mathbf{X}} = \hat{\mathbf{x}}_i)$ that is based on the following variational bound on differential entropy.

Lemma 3: Let f and g be arbitrary pdfs. If $-\int f(x) \log f(x) dx$ is finite, then $-\int f(x) \log g(x) dx$ exists and

$$-\int f(x) \log f(x) dx \leq -\int f(x) \log g(x) dx \quad (41)$$

with equality if, and only if, $f(x) = g(x)$ almost everywhere.

Proof: See [20, Lemma 8.3.1]. ■

The inequality (41) is a direct consequence of the information inequality. Lemma 3 is also reminiscent of [21, Theorem 5.1], which provides an upper bound on the mutual information between a channel input X and a channel output Y and holds for general random variables. In fact, when Y is a real-valued random variable and the conditional distribution of Y given X is absolutely continuous with respect to the Lebesgue measure, then [21, Theorem 5.1] essentially provides an upper bound on $h(Y)$ that is of the form (41).

Lemma 3 allows us to upper-bound differential entropy by replacing the true pdf f inside the logarithm by an auxiliary pdf g . In order to upper-bound the conditional differential entropy $h(\mathbf{X}|\hat{\mathbf{X}} = \hat{\mathbf{x}}_i)$, we apply Lemma 3 with the conditional pdf

$$g_{\mathbf{X}|\hat{\mathbf{X}}}(\mathbf{x}|\hat{\mathbf{x}}_i) = \begin{cases} \frac{1}{K_{i,\epsilon}}, & \mathbf{x} \in \mathcal{B}_{i,\epsilon} \\ \frac{1}{K_{i,\epsilon}} \frac{r}{\delta^{d/r}} e^{-\frac{\|\mathbf{x}-\hat{\mathbf{x}}_i\|^r}{D\delta}}, & \mathbf{x} \in \bar{\mathcal{B}}_{i,\epsilon} \end{cases} \quad (42)$$

where

$$\mathcal{B}_{i,\epsilon} \triangleq \{\mathbf{x} \in \mathcal{S}_i : \|\mathbf{x} - \hat{\mathbf{x}}_i\| \leq \epsilon\}, \quad (43a)$$

$$\bar{\mathcal{B}}_{i,\epsilon} \triangleq \{\mathbf{x} \in \mathcal{S}_i : \|\mathbf{x} - \hat{\mathbf{x}}_i\| > \epsilon\}, \quad (43b)$$

$$K_{i,\epsilon} \triangleq \Lambda_{i,\epsilon} + \frac{r}{\delta^{d/r}} \int_{\bar{\mathcal{B}}_{i,\epsilon}} e^{-\frac{\|\mathbf{x}-\hat{\mathbf{x}}_i\|^r}{D\delta}} d\mathbf{x}, \quad (43c)$$

$\Lambda_{i,\epsilon}$ denotes the Lebesgue measure of $\mathcal{B}_{i,\epsilon}$, and δ and ϵ are parameters to be specified later.

This conditional pdf of \mathbf{X} given $\hat{\mathbf{X}}$ is uniform on a set of measure $\Lambda_{i,\epsilon}$ around $\hat{\mathbf{x}}_i$ and then decays exponentially. Intuitively, if ϵ^d decays more slowly than Δ_i as D tends to zero, then with high probability \mathbf{X} lies in $\mathcal{B}_{i,\epsilon}$ and the upper bound obtained from Lemma 3 is essentially equivalent to (29) but with Δ_i replaced by $\Lambda_{i,\epsilon}$. Our choice of $g_{\mathbf{X}|\hat{\mathbf{X}}}$ for $\mathbf{x} \in \bar{\mathcal{B}}_{i,\epsilon}$ allows us to control the contribution of \mathbf{x} 's lying outside of $\mathcal{B}_{i,\epsilon}$. We next need to show that

$$\lim_{D \downarrow 0} \frac{1}{D} \sum_i p_i \Lambda_{i,\epsilon}^r \leq V_d^{r/d} \left(1 + \frac{r}{d}\right) \quad (44)$$

which corresponds to (33) generalized to arbitrary d and r , but with Δ_i replaced by $\Lambda_{i,\epsilon}$. By construction of $\mathcal{B}_{i,\epsilon}$, we have that $\sup_i \sup_{\mathbf{x} \in \mathcal{B}_{i,\epsilon}} \|\mathbf{x} - \hat{\mathbf{x}}_i\|^r \leq \epsilon^r$, so $\mathcal{B}_{i,\epsilon}$ satisfies (27) upon choosing $\epsilon^r = D/\kappa$ (for some constant

κ). The claim (44) follows therefore immediately from the steps (34)–(40). Thus, by using Lemma 3 together with (42), we can replace Δ_i (whose behavior as a function of D is unknown) by $\Lambda_{i,\epsilon}$ (whose behavior can be controlled by cleverly choosing ϵ).

Before we set out to prove Theorem 1, we first provide a number of auxiliary results that we shall need throughout the proof. The proof of Theorem 1 is then given in Section IV-B.

Lemma 4: The normalizing constant $K_{i,\epsilon}$ is upper-bounded by

$$K_{i,\epsilon} \leq \Lambda_{i,\epsilon} + dV_d D^{d/r} \Gamma\left(\frac{d}{r}, \frac{\epsilon^r}{D\delta}\right) \leq \epsilon^d V_d + dV_d D^{d/r} \Gamma\left(\frac{d}{r}\right) \quad (45)$$

where $\Gamma(\cdot, \cdot)$ denotes the upper incomplete Gamma function.

Proof: The first inequality in (45) follows from the definition of $K_{i,\epsilon}$ (43c) and by upper-bounding the integral on the RHS of (43c). Indeed, since $\bar{\mathcal{B}}_{i,\epsilon} \subseteq \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x} - \hat{\mathbf{x}}_i\| > \epsilon\}$,

$$\begin{aligned} \frac{r}{\delta^{d/r}} \int_{\bar{\mathcal{B}}_{i,\epsilon}} e^{-\frac{\|\mathbf{x} - \hat{\mathbf{x}}_i\|^r}{D\delta}} d\mathbf{x} &\leq \frac{r}{\delta^{d/r}} \int_{\|\mathbf{x} - \hat{\mathbf{x}}_i\| > \epsilon} e^{-\frac{\|\mathbf{x} - \hat{\mathbf{x}}_i\|^r}{D\delta}} d\mathbf{x} \\ &= dV_d \frac{r}{\delta^{d/r}} \int_{\rho > \epsilon} \rho^{d-1} e^{-\frac{\rho^r}{D\delta}} d\rho \\ &= dV_d D^{d/r} \int_{\xi > \frac{\epsilon^r}{D\delta}} \xi^{d/r-1} e^{-\xi} d\xi \\ &= dV_d D^{d/r} \Gamma\left(\frac{d}{r}, \frac{\epsilon^r}{D\delta}\right) \end{aligned} \quad (46)$$

where the second step follows by writing $\mathbf{x} - \hat{\mathbf{x}}_i$ in polar coordinates and by using that the surface area of the d -dimensional ball of radius $\rho = \|\mathbf{x} - \hat{\mathbf{x}}_i\|$ is $dV_d \rho^{d-1}$ (see, e.g., [16, Eq. (10)]), and the third step follows by the change of variable $\xi = \rho^r/(D\delta)$.

The second inequality in (45) follows by upper-bounding (see, e.g., [16, Eq. (7)])

$$\Lambda_{i,\epsilon} \leq \int_{\|\mathbf{x} - \hat{\mathbf{x}}_i\| \leq \epsilon} d\mathbf{x} = \epsilon^d V_d \quad (47)$$

and $\Gamma(d/r, x) \leq \Gamma(d/r)$, $x \geq 0$. ■

Lemma 5: The set $\bar{\mathcal{B}}_{i,\epsilon}$ satisfies

$$\sum_i \Pr(\mathbf{X} \in \bar{\mathcal{B}}_{i,\epsilon}) \leq \frac{D}{\epsilon^r} \quad (48a)$$

$$\sum_i \mathbb{E}[\|\mathbf{X} - \hat{\mathbf{x}}_i\|^r \mathbf{1}_{\{\mathbf{X} \in \bar{\mathcal{B}}_{i,\epsilon}\}}] \leq D. \quad (48b)$$

Proof: We first prove (48a). By the distortion constraint (3), and since $\bar{\mathcal{B}}_{i,\epsilon} \subseteq \mathcal{S}_i$ and $\|\mathbf{x} - \hat{\mathbf{x}}_i\| > \epsilon$ for $\mathbf{x} \in \bar{\mathcal{B}}_{i,\epsilon}$, we have

$$\begin{aligned} D &\geq \sum_i \int_{\mathcal{S}_i} f_{\mathbf{X}}(\mathbf{x}) \|\mathbf{x} - \hat{\mathbf{x}}_i\|^r d\mathbf{x} \\ &\geq \sum_i \int_{\bar{\mathcal{B}}_{i,\epsilon}} f_{\mathbf{X}}(\mathbf{x}) \|\mathbf{x} - \hat{\mathbf{x}}_i\|^r d\mathbf{x} \\ &\geq \sum_i \int_{\bar{\mathcal{B}}_{i,\epsilon}} f_{\mathbf{X}}(\mathbf{x}) \epsilon^r d\mathbf{x}. \end{aligned} \quad (49)$$

Using that ϵ neither depends on i nor on \mathbf{x} , (48a) follows by dividing both sides of (49) by ϵ^r .

To prove (48b) we use again the distortion constraint (3) and that $\bar{\mathcal{B}}_{i,\epsilon} \subseteq \mathcal{S}_i$ to obtain

$$\begin{aligned} \sum_i \mathbb{E} [\|\mathbf{X} - \hat{\mathbf{x}}_i\|^r \mathbb{1} \{\mathbf{X} \in \bar{\mathcal{B}}_{i,\epsilon}\}] &= \sum_i \int_{\bar{\mathcal{B}}_{i,\epsilon}} f_{\mathbf{X}}(\mathbf{x}) \|\mathbf{x} - \hat{\mathbf{x}}_i\|^r d\mathbf{x} \\ &\leq \mathbb{E} [\|\mathbf{X} - \hat{\mathbf{X}}\|^r] \\ &\leq D. \end{aligned} \quad (50)$$

■

B. Proof of Theorem 1

Expanding $I(\mathbf{X}; \hat{\mathbf{X}})$ as $h(\mathbf{X}) - h(\mathbf{X}|\hat{\mathbf{X}})$, we obtain from (5) and (21) that the excess rate can be expressed as

$$R_{r,d} = \lim_{D \downarrow 0} \left\{ \frac{d}{r} \log D + \frac{d}{r} \log \left(\frac{r}{d} (V_d \Gamma(1 + d/r))^{r/d} e \right) - \sup_{q(\cdot)} h(\mathbf{X}|\hat{\mathbf{X}}) \right\}. \quad (51)$$

To derive the lower bound (15) given in Theorem 1, it remains to show that

$$\overline{\lim}_{D \downarrow 0} \left\{ \sup_{q(\cdot)} h(\mathbf{X}|\hat{\mathbf{X}}) - \frac{d}{r} \log D \right\} \leq \frac{d}{r} \log(V_d^{r/d} (1 + r/d)). \quad (52)$$

To this end, we upper-bound the conditional differential entropy $h(\mathbf{X}|\hat{\mathbf{X}})$ using Lemma 3 together with (42). This yields for every $\hat{\mathbf{X}} = \hat{\mathbf{x}}_i$

$$\begin{aligned} h(\mathbf{X}|\hat{\mathbf{X}} = \hat{\mathbf{x}}_i) &\leq \log K_{i,\epsilon} - \mathbb{E} \left[\log \left(\frac{r}{\delta^{d/r}} e^{-\frac{\|\mathbf{X} - \hat{\mathbf{x}}_i\|^r}{D\delta}} \right) \mathbb{1} \{\mathbf{X} \in \bar{\mathcal{B}}_{i,\epsilon}\} \mid \mathbf{X} \in \mathcal{S}_i \right] \\ &\leq \log \left(\Lambda_{i,\epsilon} + dV_d D^{d/r} \Gamma \left(\frac{d}{r}, \frac{\epsilon^r}{D\delta} \right) \right) + \left| \log \left(\frac{r}{\delta^{d/r}} \right) \right| \Pr(\mathbf{X} \in \bar{\mathcal{B}}_{i,\epsilon} \mid \mathbf{X} \in \mathcal{S}_i) \\ &\quad + \frac{1}{D\delta} \mathbb{E} [\|\mathbf{X} - \hat{\mathbf{x}}_i\|^r \mathbb{1} \{\mathbf{X} \in \bar{\mathcal{B}}_{i,\epsilon}\} \mid \mathbf{X} \in \mathcal{S}_i] \end{aligned} \quad (53)$$

where the second inequality follows from the bound on $K_{i,\epsilon}$ presented in Lemma 4 and by upper-bounding $-\log(r/\delta^{d/r}) \leq |\log(r/\delta^{d/r})|$. Averaging over $\hat{\mathbf{X}}$ then yields

$$\begin{aligned} h(\mathbf{X}|\hat{\mathbf{X}}) &\leq \sum_i p_i \log \left(\Lambda_{i,\epsilon} + dV_d D^{d/r} \Gamma \left(\frac{d}{r}, \frac{\epsilon^r}{D\delta} \right) \right) \\ &\quad + \left| \log \left(\frac{r}{\delta^{d/r}} \right) \right| \sum_i \Pr(\mathbf{X} \in \bar{\mathcal{B}}_{i,\epsilon}) + \frac{1}{D\delta} \sum_i \mathbb{E} [\|\mathbf{X} - \hat{\mathbf{x}}_i\|^r \mathbb{1} \{\mathbf{X} \in \bar{\mathcal{B}}_{i,\epsilon}\}]. \end{aligned} \quad (54)$$

By Lemma 5, this can be further upper-bounded by

$$h(\mathbf{X}|\hat{\mathbf{X}}) \leq \sum_i p_i \log \left(\Lambda_{i,\epsilon} + dV_d D^{d/r} \Gamma \left(\frac{d}{r}, \frac{\epsilon^r}{D\delta} \right) \right) + \left| \log \frac{r}{\delta^{d/r}} \right| \frac{D}{\epsilon^r} + \frac{1}{\delta}. \quad (55)$$

We next choose

$$\epsilon^r = \frac{D}{\kappa} \quad (56)$$

for some $\kappa > 0$ that we will let tend to zero at the end of the proof. For ease of exposition, we do not always make this choice explicit in the notation but write ϵ^r or D/κ depending on which is more convenient.

With this choice, the second term on the RHS of (55) becomes $\kappa |\log(r/\delta^{d/r})|$. To evaluate the first term on the RHS of (55), we express p_i as

$$p_i = \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) + \Pr(\mathbf{X} \in \bar{\mathcal{B}}_{i,\epsilon}) \quad (57)$$

and define

$$\wp_\epsilon \triangleq \sum_i \Pr(\mathbf{X} \in \bar{\mathcal{B}}_{i,\epsilon}). \quad (58)$$

By Lemma 5, we have

$$\wp_\epsilon \leq \kappa \quad (59)$$

which vanishes as we let κ tend to zero. With the above definition, and applying the second inequality in (45) (Lemma 4), we obtain for the first term on the RHS of (55) that

$$\begin{aligned} & \sum_i p_i \log \left(\Lambda_{i,\epsilon} + dV_d D^{d/r} \Gamma \left(\frac{d}{r}, \frac{\epsilon^r}{D\delta} \right) \right) \\ &= \sum_i \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \log \left(\Lambda_{i,\epsilon} + dV_d D^{d/r} \Gamma \left(\frac{d}{r}, \frac{1}{\kappa\delta} \right) \right) + \sum_i \Pr(\mathbf{X} \in \bar{\mathcal{B}}_{i,\epsilon}) \log \left(\Lambda_{i,\epsilon} + dV_d D^{d/r} \Gamma \left(\frac{d}{r}, \frac{1}{\kappa\delta} \right) \right) \\ &\leq \sum_i \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \log \left(\Lambda_{i,\epsilon} + dV_d D^{d/r} \Gamma \left(\frac{d}{r}, \frac{1}{\kappa\delta} \right) \right) + \wp_\epsilon \log \left(V_d \frac{D^{d/r}}{\kappa^{d/r}} + dV_d D^{d/r} \Gamma(d/r) \right). \end{aligned} \quad (60)$$

Using (59) and that $\sum_i \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) + \wp_\epsilon = 1$, (60) becomes

$$\begin{aligned} & \sum_i p_i \log \left(\Lambda_{i,\epsilon} + dV_d D^{d/r} \Gamma \left(\frac{d}{r}, \frac{1}{\kappa\delta} \right) \right) \\ &\leq \sum_i \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \log \left(\frac{\Lambda_{i,\epsilon}}{D^{d/r}} + dV_d \Gamma \left(\frac{d}{r}, \frac{1}{\kappa\delta} \right) \right) + \wp_\epsilon \log \left(\frac{V_d}{\kappa^{d/r}} + dV_d \Gamma(d/r) \right) + \frac{d}{r} \log D \\ &\leq \frac{d}{r} \sum_i \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \log \left(\frac{\Lambda_{i,\epsilon}}{D^{d/r}} + dV_d \Gamma \left(\frac{d}{r}, \frac{1}{\kappa\delta} \right) \right)^{r/d} + \kappa \log \left(\frac{V_d}{\kappa^{d/r}} + dV_d \Gamma(d/r) \right) + \frac{d}{r} \log D. \end{aligned} \quad (61)$$

By Jensen's inequality, the first term on the RHS of (61) is upper-bounded by

$$\begin{aligned} & \frac{d}{r} \sum_i \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \log \left(\frac{\Lambda_{i,\epsilon}}{D^{d/r}} + dV_d \Gamma \left(\frac{d}{r}, \frac{1}{\kappa\delta} \right) \right)^{r/d} \\ &\leq (1 - \wp_\epsilon) \frac{d}{r} \log \left(\frac{1}{1 - \wp_\epsilon} \sum_i \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \left[\frac{\Lambda_{i,\epsilon}}{D^{d/r}} + dV_d \Gamma \left(\frac{d}{r}, \frac{1}{\kappa\delta} \right) \right]^{r/d} \right). \end{aligned} \quad (62)$$

For $r/d < 1$, we have $(x + \alpha)^{r/d} \leq x^{r/d} + \alpha^{r/d}$ for every $x, \alpha \geq 0$; for $r/d \geq 1$, the function $x \mapsto (x^{d/r} + \alpha)^{r/d}$ is concave for every $\alpha \geq 0$. Consequently,

$$\begin{aligned} & \frac{1}{1 - \wp_\epsilon} \sum_i \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \left[\frac{\Lambda_{i,\epsilon}}{D^{d/r}} + dV_d \Gamma \left(\frac{d}{r}, \frac{1}{\kappa\delta} \right) \right]^{r/d} \\ &\leq \begin{cases} \frac{1}{1 - \wp_\epsilon} \sum_i \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{D} + d^{r/d} V_d^{r/d} \Gamma \left(\frac{d}{r}, \frac{1}{\kappa\delta} \right)^{r/d}, & r/d < 1 \\ \left[\left(\frac{1}{1 - \wp_\epsilon} \sum_i \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{D} \right)^{d/r} + dV_d \Gamma \left(\frac{d}{r}, \frac{1}{\kappa\delta} \right) \right]^{r/d}, & r/d \geq 1 \end{cases} \end{aligned} \quad (63)$$

where the upper bound for $r/d \geq 1$ follows from Jensen's inequality.

We next generalize (33), namely,

$$\lim_{D \downarrow 0} \frac{1}{D} \sum_i p_i \Delta_i^2 \leq 12 \quad (64)$$

to the d -dimensional sets $\mathcal{B}_{i,\epsilon}$ of Lebesgue measure $\Lambda_{i,\epsilon}$. To this end, we follow essentially the steps (34)–(40) in Section III with \mathcal{S}_i replaced by $\mathcal{B}_{i,\epsilon}$ and with Δ_i replaced by $\Lambda_{i,\epsilon}$. However, (39) is based on Lebesgue's

differentiation theorem, which requires that the families of sets $\mathcal{B}_{i,\epsilon}$ (parametrized by D) have *bounded eccentricity*.² Since $\mathcal{B}_{i,\epsilon}$ is the intersection of \mathcal{S}_i with the d -dimensional ball of radius ϵ centered at $\hat{\mathbf{x}}_i$, cf. (43a), and since \mathcal{S}_i is arbitrary, the sets $\mathcal{B}_{i,\epsilon}$ may not fulfill this condition. In the one-dimensional case, a sufficient condition for $\mathcal{B}_{i,\epsilon}$ having bounded eccentricity would be that, for every distortion D , the quantization regions \mathcal{S}_i are convex. This in turn can be assumed without loss of optimality, e.g., for quadratic distortion and sources with well-behaved pdfs [11]. However, for one-dimensional sources with general pdfs, or for higher-dimensional sources, assuming convex quantization regions may be too restrictive. Fortunately, the families of sets $\mathcal{B}_{i,\epsilon}$ that have not bounded eccentricity can be disregarded without affecting the final result. The inequality (33) can therefore be generalized to the case at hand without imposing any additional constraints on the quantization regions \mathcal{S}_i , $i \in \mathbb{Z}$ or the source pdf $f_{\mathbf{X}}$. The result is stated in the following lemma.

Lemma 6: Let the sets $\mathcal{B}_{i,\epsilon}$, $i \in \mathbb{Z}$ be defined in (43a), and let $\Lambda_{i,\epsilon}$, $i \in \mathbb{Z}$ denote the Lebesgue measures of these sets. Assume that $\epsilon^r = D/\kappa$. Then, for every $\kappa > 0$,

$$\overline{\lim}_{D \downarrow 0} \sup_{q(\cdot)} \sum_i \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{D} \leq V_d^{r/d} \left(1 + \frac{r}{d}\right). \quad (65)$$

Proof: See Appendix A. ■

Combining Lemma 6 with (55)–(63), and bounding $0 \leq \wp_\epsilon \leq \kappa$, we obtain that

$$\begin{aligned} \overline{\lim}_{D \downarrow 0} \left\{ \sup_{q(\cdot)} h(\mathbf{X}|\hat{\mathbf{X}}) - \frac{d}{r} \log D \right\} &\leq \frac{d}{r} \log \left(\frac{V_d^{r/d}(1+r/d)}{1-\kappa} + d^{r/d} V_d^{r/d} \Gamma\left(\frac{d}{r}, \frac{1}{\kappa\delta}\right)^{r/d} \right) \\ &\quad + \kappa \log \left(\frac{V_d}{\kappa^{d/r}} + d V_d \Gamma(d/r) \right) + \kappa \left| \log \frac{r}{\delta^{d/r}} \right| + \frac{1}{\delta}, \quad \text{for } r/d < 1 \end{aligned} \quad (66a)$$

and

$$\begin{aligned} \overline{\lim}_{D \downarrow 0} \left\{ \sup_{q(\cdot)} h(\mathbf{X}|\hat{\mathbf{X}}) - \frac{d}{r} \log D \right\} &\leq \log \left(\frac{V_d(1+r/d)^{d/r}}{(1-\kappa)^{d/r}} + d V_d \Gamma\left(\frac{d}{r}, \frac{1}{\kappa\delta}\right) \right) \\ &\quad + \kappa \log \left(\frac{V_d}{\kappa^{d/r}} + d V_d \Gamma(d/r) \right) + \kappa \left| \log \frac{r}{\delta^{d/r}} \right| + \frac{1}{\delta}, \quad \text{for } r/d \geq 1. \end{aligned} \quad (66b)$$

Using that $\lim_{\xi \rightarrow \infty} \Gamma(d/r, \xi) = 0$ and $\lim_{\xi \rightarrow 0} \xi \log(\alpha/\xi^{d/r} + \beta) = 0$ (for any $\alpha, \beta > 0$), letting $\kappa \rightarrow 0$ yields

$$\overline{\lim}_{D \downarrow 0} \left\{ \sup_{q(\cdot)} h(\mathbf{X}|\hat{\mathbf{X}}) - \frac{d}{r} \log D \right\} \leq \frac{d}{r} \log(V_d^{r/d}(1+r/d)) + \frac{1}{\delta}. \quad (67)$$

This in turn proves (52) upon letting $\delta \rightarrow \infty$ and concludes the proof of Theorem 1.

V. ASYMPTOTICALLY OPTIMAL QUANTIZERS

As mentioned at the end of Section II, in the one-dimensional case uniform quantizers with cells of length $2(1+r)^{1/r} D^{1/r}$ achieve the asymptotic excess rate $R_{r,1}$. Hence, uniform quantizers are asymptotically optimal as the allowed distortion tends to zero. One may wonder whether every sequence of quantizers achieving $R_{r,1}$ must converge to a uniform quantizer as $D \rightarrow 0$, or whether uniform quantizers are merely a convenient choice

²A family \mathcal{F} of sets is said to have bounded eccentricity if there exists a constant $c > 0$ such that for every $\mathcal{S} \in \mathcal{F}$ the Lebesgue measure of \mathcal{S} is not smaller than c times the volume of the smallest ball containing \mathcal{S} .

and other quantizers with vanishing cells are also asymptotically optimal. In this section, we partially address this question by presenting in Theorem 7 a necessary condition for the asymptotic optimality of a sequence of quantizers (parametrized by D). We then apply this condition to the family of almost-regular quantizers.

Theorem 7: Suppose the sequence of quantizers $q(\cdot)$ (parametrized by D) with quantization regions \mathcal{S}_i , $i \in \mathbb{Z}$ satisfying the distortion constraint $\mathbb{E}[|X - q(X)|^r] \leq D$ achieves the asymptotic excess distortion

$$\lim_{D \downarrow 0} \{H(q(X)) - R(D)\} = \frac{1}{r} \log \left(\frac{\Gamma(1 + 1/r)^r e}{1 + 1/r} \right). \quad (68)$$

Then,

$$\lim_{\rho \rightarrow \infty} \overline{\lim}_{D \downarrow 0} \sum_i \Pr(X \in \mathcal{S}_i) \mathbb{1} \left\{ \left| \frac{\Lambda_{i, \rho D^{1/r}}^r}{D} - 2^r(1 + r) \right| \leq \vartheta \right\} = 1, \quad \text{for every } \vartheta > 0. \quad (69)$$

Here, $\Lambda_{i, \rho D^{1/r}}$ denotes the Lebesgue measure of $\mathcal{B}_{i, \epsilon}$ in (43a) for $\epsilon = \rho D^{1/r}$.

Proof: This result is a direct consequence of Jensen's inequality applied in (62) in the proof of Theorem 1. See Appendix B for a detailed proof. ■

If we interpret the quantizer as a random variable that takes on the value \mathcal{S}_i with probability $\Pr(X \in \mathcal{S}_i)$, then Theorem 7 can be paraphrased as follows: “A sequence of quantizer achieves the asymptotic excess distortion $R_{r,1}$ only if $\Lambda_{i, \rho D^{1/r}}$ converges in probability to $2(1 + r)^{1/r} D^{1/r}$ as $D \rightarrow 0$ and $\rho \rightarrow \infty$.”

A quantizer $q(\cdot)$ is said to be *almost regular* if there exists a set $\bar{\mathcal{S}} \subset \mathcal{X}$ of Lebesgue measure zero such that on $\mathcal{X} \setminus \bar{\mathcal{S}}$ the quantization regions are intervals containing the reconstruction value [11]. (For all $x \in \bar{\mathcal{S}}$, we can define $q(x)$ in an arbitrary manner without changing the entropy and distortion of $q(\cdot)$.) In other words, an almost-regular quantizer $q(\cdot)$ can be written as

$$q(x) = \sum_i c_i \mathbb{1} \{a_i \leq x < b_i\}, \quad \text{for } x \in \mathcal{X} \setminus \bar{\mathcal{S}} \quad (70a)$$

$$q(x) = \sum_i \bar{x}_i \mathbb{1} \{x \in \bar{\mathcal{S}}_i\}, \quad \text{for } x \in \bar{\mathcal{S}} \quad (70b)$$

where $a_i \leq c_i \leq b_i$, and where \bar{x}_i and $\bar{\mathcal{S}}_i$ are arbitrary.

For almost-regular quantizers, condition (69) in Theorem 7 can be simplified as follows. Firstly, since the source has a pdf and $\bar{\mathcal{S}}$ has measure zero,

$$\sum_i \Pr(X \in \mathcal{S}_i \cap \bar{\mathcal{S}}) = 0. \quad (71)$$

Secondly, for any quantization region $[a_i, b_i) \subseteq \mathcal{X} \setminus \bar{\mathcal{S}}$ and reconstruction value $c_i \in [a_i, b_i)$, we have

$$\min \{\Delta_i^r, \rho^r D\} \leq \Lambda_{i, \rho D^{1/r}}^r \leq \Delta_i^r \quad (72)$$

where $\Delta_i = b_i - a_i$. Consequently,

$$\mathbb{1} \left\{ \left| \frac{\Lambda_{i, \rho D^{1/r}}^r}{D} - 2^r(1 + r) \right| \leq \vartheta \right\} = \mathbb{1} \left\{ \left| \frac{\Delta_i}{D} - 2^r(1 + r) \right| \leq \vartheta \right\}, \quad \rho \geq (2^r(1 + r))^{1/r}. \quad (73)$$

We thus have the following result:

Corollary 8: Suppose the sequence of almost-regular quantizers $q(\cdot)$ (parametrized by D) with quantization regions \mathcal{S}_i , $i \in \mathbb{Z}$ satisfying the distortion constraint $\mathbb{E}[|X - q(X)|^r] \leq D$ achieves the asymptotic excess distortion

$$\lim_{D \downarrow 0} \{H(q(X)) - R(D)\} = \frac{1}{r} \log \left(\frac{\Gamma(1 + 1/r)^r e}{1 + 1/r} \right). \quad (74)$$

Then,

$$\overline{\lim}_{D \downarrow 0} \sum_i \Pr(X \in \mathcal{S}_i) \mathbb{1} \left\{ \left| \frac{\Delta_i}{D} - 2^r(1+r) \right| \leq \vartheta \right\} = 1, \quad \text{for every } \vartheta > 0. \quad (75)$$

Here, Δ_i denotes the Lebesgue measure of \mathcal{S}_i .

Again, interpreting the quantizer as a random variable that takes on the value \mathcal{S}_i with probability $\Pr(X \in \mathcal{S}_i)$, Corollary 8 can be paraphrased as “any sequence of almost-regular quantizers achieving $\bar{R}_{r,1}$ must converge in probability to a uniform quantizer as $D \rightarrow 0$.”

VI. BALLS VERSUS TESSELLATING POLYTOPES

The lower bound (15) on the excess rate presented in Theorem 1 hinges on the fact that the distortion over the quantization region \mathcal{S}_i , i.e., $\int_{\mathcal{S}_i} \|\mathbf{x} - \hat{\mathbf{x}}\|^r d\mathbf{x}$, is lower-bounded by the distortion over a ball around $\hat{\mathbf{x}}_i$ with the same volume (cf. (84) in the proof of Theorem 1 with $\mathcal{B}_{i,\epsilon}$ replaced by \mathcal{S}_i and with $\Lambda_{i,\epsilon}$ replaced by Δ_i). Since the one-dimensional ball is an interval and, hence, tessellates \mathbb{R} , it follows that for scalar sources the lower bound (15) is achieved by a tessellating quantizer, so in this case it is tight. However, it is expected that this is no longer true for multi-dimensional sources, since in general balls do not tessellate the space. In fact, it is unclear whether there exists any (possibly non-tessellating) vector quantizer that achieves (15) for multi-dimensional sources.

To assess the tightness of the obtained lower bound, we compare it numerically with the excess rates achievable by several lattice quantizers. To this end, we use Linder and Zeger’s upper bound for tessellating quantizers (23) together with the normalized second moments $\ell(\mathcal{P})$ of various lattice quantizers tabulated in [22, Table I]. In order to better compare our results with previous works, in this section we consider the *excess rate per dimension*, defined as $\bar{R}_{r,d} \triangleq R_{r,d}/d$. The excess rate per dimension is relevant, for example, in the analysis of quantization schemes that buffer d consecutive symbols of a one-dimensional memoryless source and then quantize them using a d -dimensional vector quantizer.

For the sake of simplicity, we only consider quadratic distortion and the Euclidean norm. In this case, the lower bound (15) becomes

$$\bar{R}_{2,d} \geq \frac{1}{2} \log \left(2\pi e \frac{\Gamma(1 + d/2)^{2/d}}{\pi(2+d)} \right). \quad (76)$$

Furthermore, the upper bound corresponding to tessellating quantizers (23) becomes

$$\bar{R}_{2,d} \leq \frac{1}{2} \log \left(2\pi e \frac{1}{d} \inf_{\mathcal{P}} \ell(\mathcal{P}) \right). \quad (77)$$

Another upper bound on $\bar{R}_{2,d}$ follows from an upper bound on $b_{r,d}$ in Zador’s theorem (9) that was presented in [6]. This upper bound is based on random coding arguments and yields for quadratic distortion and the Euclidean norm

$$\bar{R}_{2,d} \leq \frac{1}{2} \log \left(2\pi e \frac{\Gamma(1 + 2/d) \Gamma(1 + d/2)^{2/d}}{\pi d} \right). \quad (78)$$

The bound (78) demonstrates that $\bar{R}_{2,d}$ vanishes as d tends to infinity. This is perhaps not very surprising, since the rate-distortion function $R(D)$ is essentially achieved by a vector quantizer whose dimension tends to infinity.

In Figure 1, we depict the bounds (76) and (78) as a function of the dimension d . We further show several achievability results based on lattice quantizers (77). The normalized second moments $\ell(\mathcal{P})$ corresponding to

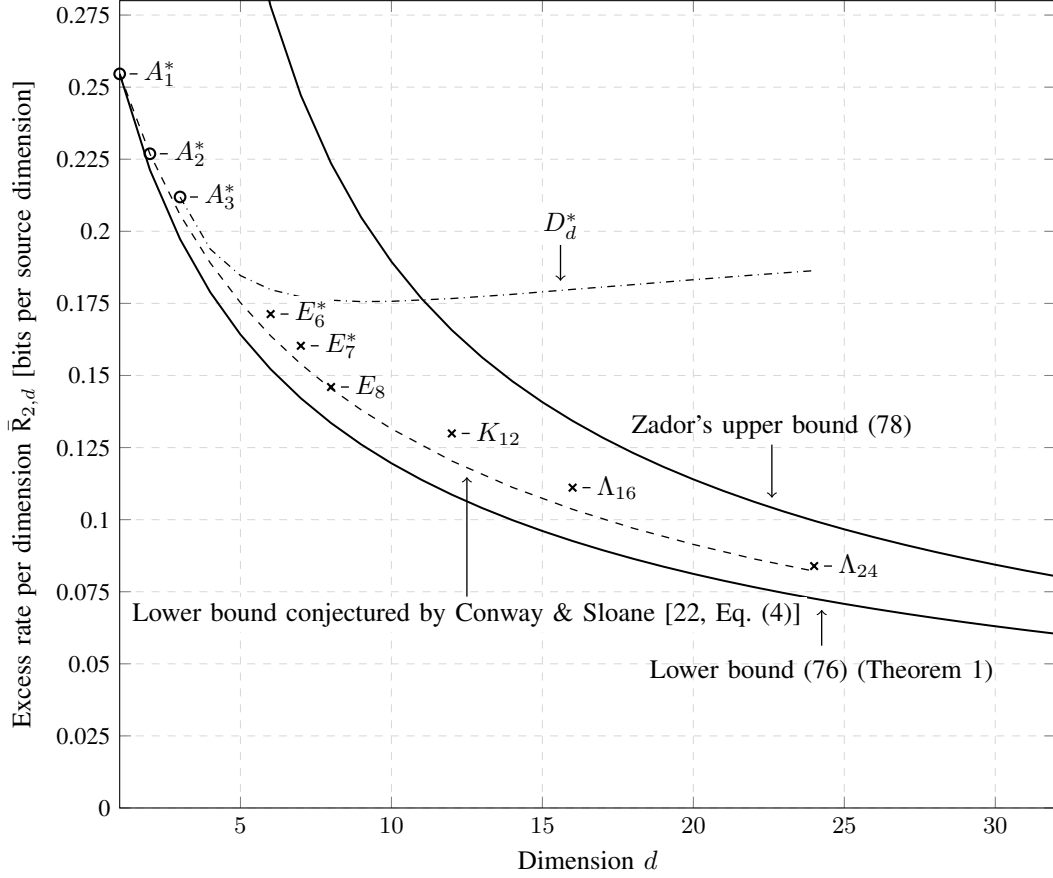


Figure 1: Bounds on the excess rate per dimension $\bar{R}_{2,d}$ (in bits per source dimension) of a d -dimensional vector quantizer. The excess rate per dimension attained by lattice quantizers was obtained by applying to (77) the normalized second moments tabulated in [22, Table I].

these lattice quantizers were tabulated by Conway and Sloane in [22, Table I]. In fact, Figure 1 is equivalent to [22, Figure 1] with the only difference that here we plot the excess rate per dimension whereas Conway and Sloane plot the normalized second moment. Specifically, we include the excess rates per dimension incurred by a (one-dimensional) uniform quantizer, by a (two-dimensional) hexagonal quantizer, and by the three-dimensional tessellating quantizer whose regions are cuboctahedrons. These quantizers correspond to the so-called Voronoi lattices of the first type A_1^* (the integers), A_2^* (the two-dimensional hexagonal lattice), and A_3^* (the body-centered cubic lattice). For $d \geq 3$, we further include the excess rates per dimension attained by the D_d^* lattices. Labeled with cross markers, we show the excess rates per dimension corresponding to the lattices E_6^* , E_7^* , the Gosset lattice E_8 , the Coxeter-Todd lattice K_{12} , the Barnes-Wall lattice Λ_{16} , and the Leech lattice Λ_{24} . We refer to [22] and references therein for further details.

Finally, we compare the obtained bounds with a conjectured lower bound by Conway and Sloane [22, Eq. (4)] that follows by computing the distortion attained by a set of reconstruction points located at the vertices of a d -dimensional tetrahedron. Note that this bound was computed for *fixed-rate quantizers*, i.e., for quantizers that have

a finite number M of quantization regions and whose rate is defined as $\log M$. While the excess rate achievable by a fixed-rate quantizer can also be achieved by an entropy-constrained quantizer, the converse is not necessarily true. It is thus *prima facie* unclear whether Conway and Sloane's conjectured lower bound would also apply to entropy-constrained quantizers. Nevertheless, we decided to include it here since it is remarkably close to the excess rates per dimension corresponding to lattices E_8 and Λ_{24} .

As mentioned above, the excess rate per dimension vanishes as d tends to infinity. However, as illustrated by Figure 1, it decays slowly: for example, for a 10-dimensional vector quantizer we still have

$$\bar{R}_{2,10} \geq \frac{1}{2} \log_2 \left(\frac{e\Gamma(6)^{1/5}}{6} \right) \approx 0.1196 \text{ bits per source dimension} \quad (79)$$

which is, arguably, not much smaller than the excess rate per dimension of the (one-dimensional) uniform quantizer

$$\bar{R}_{2,1} = \frac{1}{2} \log_2(\pi e/6) \approx 0.2546 \text{ bits per source dimension.}$$

(Here $\log_2(\cdot)$ denotes the binary logarithm.) In general, the bounds on $\bar{R}_{2,d}$ given in (76) and (78) are of the order $\Theta(\log d/d)$.

Observe that for multi-dimensional sources the gap between the lower bound (76) and the excess rate per dimension achievable with lattice quantizers is substantial. This gap is partly due to the fact that, in order to derive the lower bound (15), we lower-bounded the distortion over the quantization region \mathcal{S}_i by that over a ball with the same volume, cf. (84). To obtain a tighter lower bound, we may need a more accurate approximation of this distortion that, like the conjectured bound by Conway and Sloane, takes the geometry of the optimal quantization regions into account.

VII. CONCLUSIONS

The nonnegativity of relative entropy implies that the differential entropy of a random variable X with pdf f is upper-bounded by $-\mathbb{E}[\log g(X)]$ for any arbitrary pdf g . Using this inequality with a cleverly chosen g , we derived a lower bound on the asymptotic excess rate of entropy-constrained scalar quantization. Specialized to the one-dimensional case and quadratic distortion, this bound coincides with the excess rate obtained by Gish and Pierce in [4], and by Gray *et al.* in [10] particularized for scalar quantizers. The proposed derivation thus recovers the well-known result that uniform quantizers are asymptotically optimal as the allowed distortion vanishes.

Our result holds for any d -dimensional memoryless source \mathbf{X} that satisfies $|h(\mathbf{X})| < \infty$ and $H(\lfloor \mathbf{X} \rfloor) < \infty$. The presented proof is thus as general as the proof by Gray *et al.*, and it is more general than the proof by Gish and Pierce. In fact, it has recently been shown that these conditions are necessary and sufficient for the Shannon lower bound to be asymptotically tight for vanishing distortion, and that $H(\lfloor \mathbf{X} \rfloor) < \infty$ is a necessary and sufficient condition for the rate-distortion function to be finite [9]. Our result thus holds for the most general conditions that can be imposed in the analysis of high-resolution quantizers.

The derivation of the lower bound reveals a necessary condition for a sequence of quantizers (parametrized by D) to achieve the asymptotic excess rate. Specifically, we demonstrated for scalar sources that the intersection of the quantization region \mathcal{S}_i with the interval $[\hat{x}_i - \rho D^{1/r}, \hat{x}_i + \rho D^{1/r}]$ must have a Lebesgue measure that converges in

probability to $2(1+r)^{1/r}D^{1/r}$ as $D \rightarrow 0$ and $\rho \rightarrow \infty$. This implies that any sequence of almost-regular quantizers achieving the asymptotic excess rate must converge in probability to a uniform quantizer as $D \rightarrow 0$. Since almost-regular quantizers achieve $D_{r,1}(R)$ when $r \geq 1$, this in turn suggests that asymptotically-optimal quantizers must essentially be uniform.

While the presented bound is tight for scalar sources, it is unclear whether the same is true for multi-dimensional sources. Indeed, its derivation hinges on the fact that the distortion over the quantization region \mathcal{S}_i is lower-bounded by the distortion over a ball around $\hat{\mathbf{x}}_i$ with the same volume, cf. (84). Since the one-dimensional ball is an interval and, hence, tessellates \mathbb{R} , it follows that for one-dimensional sources the converse bound (15) is achieved by a tessellating quantizer (which in this case is the uniform quantizer). However, it is expected that this is no longer true for multi-dimensional sources, since in general balls do not tessellate the space. It is yet unclear whether there exists any (possibly non-tessellating) vector quantizer that achieves our converse bound for multi-dimensional sources.

APPENDIX A

PROOF OF LEMMA 6

To prove Lemma 6, we first fix an arbitrary constant $\eta > 0$ and divide the indices i according to whether $\Lambda_{i,\epsilon} \geq \eta V_d \epsilon^d$ or not. Specifically, let

$$\mathcal{I} \triangleq \{i \in \mathbb{Z} : \Lambda_{i,\epsilon} \geq \eta V_d \epsilon^d\} \quad (80)$$

and divide the sum on the LHS of (65) into

$$\sum_i \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{D} = \sum_{i \in \mathcal{I}} \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{D} + \sum_{i \in \mathcal{I}^c} \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{D} \quad (81)$$

where \mathcal{I}^c denotes the complement of \mathcal{I} . For every $i \in \mathcal{I}^c$ we have $\Lambda_{i,\epsilon} < \eta V_d \epsilon^d$, so the second sum on the RHS of (81) can be upper-bounded as

$$\begin{aligned} \sum_{i \in \mathcal{I}^c} \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{D} &\leq \eta^{r/d} V_d^{r/d} \frac{\epsilon^r}{D} \sum_{i \in \mathcal{I}^c} \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \\ &\leq \eta^{r/d} \frac{V_d^{r/d}}{\kappa} \end{aligned} \quad (82)$$

where the second step follows because $\epsilon^r = D/\kappa$ and because, by definition, the sets $\mathcal{B}_{i,\epsilon}$ are disjoint, so the sum of the probabilities $\Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon})$ is equal to the probability of $\cup_{i \in \mathcal{I}^c} \mathcal{B}_{i,\epsilon}$, which is upper-bounded by 1.

To upper-bound the first sum on the RHS of (81), we begin by lower-bounding $\mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}\|^r]$ as

$$\begin{aligned} \mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}\|^r] &= \sum_i \int_{\mathcal{S}_i} f_{\mathbf{X}}(\mathbf{x}) \|\mathbf{x} - \hat{\mathbf{x}}_i\|^r d\mathbf{x} \\ &\geq \sum_{i \in \mathcal{I}} \int_{\mathcal{B}_{i,\epsilon}} f_{\mathbf{X}}(\mathbf{x}) \|\mathbf{x} - \hat{\mathbf{x}}_i\|^r d\mathbf{x} \\ &= \sum_{i \in \mathcal{I}} \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \frac{1}{\Lambda_{i,\epsilon}} \int_{\mathcal{B}_{i,\epsilon}} \|\mathbf{x} - \hat{\mathbf{x}}_i\|^r d\mathbf{x} \\ &\quad - \sum_{i \in \mathcal{I}} \int_{\mathcal{B}_{i,\epsilon}} \left[\frac{1}{\Lambda_{i,\epsilon}} \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) - f_{\mathbf{X}}(\mathbf{x}) \right] \|\mathbf{x} - \hat{\mathbf{x}}_i\|^r d\mathbf{x}. \end{aligned} \quad (83)$$

The region $\mathcal{B}_{i,\epsilon}$ of volume $\Lambda_{i,\epsilon}$ that minimizes $\int_{\mathcal{B}_{i,\epsilon}} \|\mathbf{x} - \hat{\mathbf{x}}\|^r d\mathbf{x}$ is a ball around $\hat{\mathbf{x}}$. We thus have [16, Section III]

$$\frac{1}{\Lambda_{i,\epsilon}} \int_{\mathcal{B}_{i,\epsilon}} \|\mathbf{x} - \hat{\mathbf{x}}_i\|^r d\mathbf{x} \geq \frac{d}{d+r} \frac{\Lambda_{i,\epsilon}^{r/d}}{V_d^{r/d}} \quad (84)$$

which yields for the first term on the RHS of (83)

$$\sum_{i \in \mathcal{I}} \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \frac{1}{\Lambda_{i,\epsilon}} \int_{\mathcal{B}_{i,\epsilon}} \|\mathbf{x} - \hat{\mathbf{x}}_i\|^r d\mathbf{x} \geq \sum_{i \in \mathcal{I}} \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{V_d^{r/d}(1+r/d)}. \quad (85)$$

Multiplying both sides of (83) by $V_d^{r/d}(1+r/d)/D$, applying (85) to (83), and using that $\mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}\|^r] \leq D$, we obtain

$$\sum_{i \in \mathcal{I}} \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{D} \leq V_d^{r/d} \left(1 + \frac{r}{d}\right) \left(1 + \frac{1}{D} \sum_{i \in \mathcal{I}} \int_{\mathcal{B}_{i,\epsilon}} \left[\frac{1}{\Lambda_{i,\epsilon}} \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) - f_{\mathbf{X}}(\mathbf{x}) \right] \|\mathbf{x} - \hat{\mathbf{x}}_i\|^r d\mathbf{x} \right). \quad (86)$$

We next introduce the pdf

$$\begin{aligned} f_{\mathbf{X}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) &\triangleq \sum_{i \in \mathcal{I}} \frac{1}{\Lambda_{i,\epsilon}} \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \mathbb{1}\{\mathbf{x} \in \mathcal{B}_{i,\epsilon}\} \\ &\quad + f_{\mathbf{X}}(\mathbf{x}) \left[\sum_{i \in \mathcal{I}} \mathbb{1}\{\mathbf{x} \in \bar{\mathcal{B}}_{i,\epsilon}\} + \sum_{i \in \mathcal{I}^c} \mathbb{1}\{\mathbf{x} \in \mathcal{S}_i\} \right], \quad \mathbf{x} \in \mathbb{R}^d \end{aligned} \quad (87)$$

which allows us to write

$$\sum_{i \in \mathcal{I}} \int_{\mathcal{B}_{i,\epsilon}} \left[\frac{1}{\Lambda_{i,\epsilon}} \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) - f_{\mathbf{X}}(\mathbf{x}) \right] \|\mathbf{x} - \hat{\mathbf{x}}_i\|^r d\mathbf{x} = \sum_i \int_{\mathcal{S}_i} \left[f_{\mathbf{X}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) - f_{\mathbf{X}}(\mathbf{x}) \right] \|\mathbf{x} - \hat{\mathbf{x}}_i\|^r d\mathbf{x}. \quad (88)$$

Since $\|\mathbf{x} - \hat{\mathbf{x}}_i\|^r \leq \epsilon^r = D/\kappa$ for $\mathbf{x} \in \mathcal{B}_{i,\epsilon}$, $i \in \mathcal{I}$ and $f_{\mathbf{X}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) - f_{\mathbf{X}}(\mathbf{x}) = 0$ otherwise, we have

$$\left| \sum_i \int_{\mathcal{S}_i} \left[f_{\mathbf{X}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) - f_{\mathbf{X}}(\mathbf{x}) \right] \|\mathbf{x} - \hat{\mathbf{x}}_i\|^r d\mathbf{x} \right| \leq \frac{D}{\kappa} \int \left| f_{\mathbf{X}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) - f_{\mathbf{X}}(\mathbf{x}) \right| d\mathbf{x}. \quad (89)$$

Combining this upper bound with (81), (82), and (86), we obtain

$$\sum_i \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{D} \leq V_d^{r/d} \left(1 + \frac{r}{d}\right) \left(1 + \frac{1}{\kappa} \int \left| f_{\mathbf{X}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) - f_{\mathbf{X}}(\mathbf{x}) \right| d\mathbf{x} \right) + \eta^{r/d} \frac{V_d^{r/d}}{\kappa}. \quad (90)$$

We next show that, for every $\eta > 0$,

$$\limsup_{D \downarrow 0} \int_{q(\cdot)} \left| f_{\mathbf{X}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) - f_{\mathbf{X}}(\mathbf{x}) \right| d\mathbf{x} = 0. \quad (91)$$

(Note that $f_{\mathbf{X}}^{(\Lambda)}$ depends on $q(\cdot)$ and D via $\mathcal{B}_{i,\epsilon}$, $i \in \mathbb{Z}$.) It then follows that

$$\overline{\lim}_{D \downarrow 0} \sup_{q(\cdot)} \sum_i \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{D} \leq V_d^{r/d} \left(1 + \frac{r}{d}\right) + \eta^{r/d} \frac{V_d^{r/d}}{\kappa} \quad (92)$$

which proves Lemma 6 upon letting η tend to zero from above.

It thus remains to prove (91). By definition, $f_{\mathbf{X}}^{(\Lambda)}$ differs from $f_{\mathbf{X}}$ only when $\mathbf{x} \in \mathcal{B}_{i,\epsilon}$, $i \in \mathcal{I}$. Since the family of sets $\mathcal{B}_{i,\epsilon}$, $i \in \mathcal{I}$ (parametrized by D) has bounded eccentricity, it follows from Lebesgue's differentiation theorem that $f_{\mathbf{X}}^{(\Lambda)}$ converges to $f_{\mathbf{X}}$ almost everywhere as D (and hence also ϵ) tends to zero, which by Scheffe's lemma then implies (91). However, compared to the standard setting under which Lebesgue's differentiation theorem is proven, our setting is slightly more complicated, since as D tends to zero not only the diameters of the sets $\mathcal{B}_{i,\epsilon}$

decay, but also their locations in \mathbb{R}^d may change. For completeness, we therefore provide all the steps, even though they follow closely the standard proof of the Lebesgue differentiation theorem.

We first note that the integral in (91) is nonnegative and bounded, so its supremum is finite and for every $\nu > 0$ there exists a sequence of quantizers (parametrized by D) such that

$$\overline{\lim}_{D \downarrow 0} \int \left| f_{\mathbf{X}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) - f_{\mathbf{X}}(\mathbf{x}) \right| d\mathbf{x} \geq \overline{\lim}_{D \downarrow 0} \sup_{q(\cdot)} \int \left| f_{\mathbf{X}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) - f_{\mathbf{X}}(\mathbf{x}) \right| d\mathbf{x} - \nu. \quad (93)$$

Since $\nu > 0$ is arbitrary, it follows that, in order to prove (91), it suffices to show that for any sequence of quantizers (parametrized by D)

$$\lim_{D \downarrow 0} \int \left| f_{\mathbf{X}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) - f_{\mathbf{X}}(\mathbf{x}) \right| d\mathbf{x} = 0. \quad (94)$$

Specifically, we shall show that for any sequence of quantizers (parametrized by D)

$$\lambda \left(\left\{ \mathbf{x} \in \mathbb{R}^d : \overline{\lim}_{D \downarrow 0} \left| f_{\mathbf{X}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) - f_{\mathbf{X}}(\mathbf{x}) \right| > 2\xi \right\} \right) = 0, \quad \text{for every } \xi > 0 \quad (95)$$

where $\lambda(\cdot)$ denotes the Lebesgue measure on \mathbb{R}^d . It then follows that $f_{\mathbf{X}}^{(\Lambda)}$ converges to $f_{\mathbf{X}}$ almost everywhere as $D \rightarrow 0$ since

$$\left\{ \mathbf{x} \in \mathbb{R}^d : \overline{\lim}_{D \downarrow 0} \left| f_{\mathbf{X}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) - f_{\mathbf{X}}(\mathbf{x}) \right| > 0 \right\} = \bigcup_{\ell=1}^{\infty} \left\{ \mathbf{x} \in \mathbb{R}^d : \overline{\lim}_{D \downarrow 0} \left| f_{\mathbf{X}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) - f_{\mathbf{X}}(\mathbf{x}) \right| > \frac{1}{\ell} \right\} \quad (96)$$

and the countable union of sets of measure zero has measure zero. By Scheffe's lemma, almost everywhere convergence of $f_{\mathbf{X}}^{(\Lambda)}$ to $f_{\mathbf{X}}$ implies (94), which together with (93) proves the desired result (91).

We thus set out to prove (95). By the definition of $f_{\mathbf{X}}^{(\Lambda)}$ and the triangle inequality,

$$\left| f_{\mathbf{X}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) - f_{\mathbf{X}}(\mathbf{x}) \right| \leq \sum_{i \in \mathcal{I}} \left| \frac{1}{\Lambda_{i,\epsilon}} \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) - f_{\mathbf{X}}(\mathbf{x}) \right| \mathbb{1}_{\{\mathbf{x} \in \mathcal{B}_{i,\epsilon}\}}. \quad (97)$$

We next approximate $\frac{1}{\Lambda_{i,\epsilon}} \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon})$ by replacing $f_{\mathbf{X}}$ by a continuous function g . Indeed, since $f_{\mathbf{X}}$ is integrable, for every $\varepsilon > 0$ there exists a continuous function g such that [23, Theorem 2.4.14, p. 92]

$$\int |f_{\mathbf{X}}(\mathbf{x}) - g(\mathbf{x})| d\mathbf{x} \leq \varepsilon. \quad (98)$$

It then follows that, for every $\mathbf{x} \in \mathcal{B}_{i,\epsilon}$,

$$\begin{aligned} \left| \frac{1}{\Lambda_{i,\epsilon}} \Pr(\mathbf{X} \in \mathcal{B}_{i,\epsilon}) - f_{\mathbf{X}}(\mathbf{x}) \right| &\leq \left| \frac{1}{\Lambda_{i,\epsilon}} \int_{\mathcal{B}_{i,\epsilon}} g(\mathbf{y}) d\mathbf{y} - g(\mathbf{x}) \right| \\ &\quad + \frac{1}{\Lambda_{i,\epsilon}} \int_{\mathcal{B}_{i,\epsilon}} |f_{\mathbf{X}}(\mathbf{y}) - g(\mathbf{y})| d\mathbf{y} + |f_{\mathbf{X}}(\mathbf{x}) - g(\mathbf{x})|. \end{aligned} \quad (99)$$

Let $\mathcal{B}(\mathbf{c}, \rho) \triangleq \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x} - \mathbf{c}\| \leq \rho\}$ denote the d -dimensional ball of radius ρ centered at \mathbf{c} . Note that $\lambda(\mathcal{B}(\hat{\mathbf{x}}_i, \epsilon)) = V_d \epsilon^d$. For every $\mathbf{x} \in \mathcal{B}_{i,\epsilon}$ and $i \in \mathcal{I}$, the second term on the RHS of (99) can be upper-bounded by

$$\begin{aligned} \frac{1}{\Lambda_{i,\epsilon}} \int_{\mathcal{B}_{i,\epsilon}} |f_{\mathbf{X}}(\mathbf{y}) - g(\mathbf{y})| d\mathbf{y} &\leq \frac{1}{\eta \lambda(\mathcal{B}(\hat{\mathbf{x}}_i, \epsilon))} \int_{\mathcal{B}(\hat{\mathbf{x}}_i, \epsilon)} |f_{\mathbf{X}}(\mathbf{y}) - g(\mathbf{y})| d\mathbf{y} \\ &\leq \frac{2^d}{\eta} \frac{1}{\lambda(\mathcal{B}(\mathbf{x}, 2\epsilon))} \int_{\mathcal{B}(\mathbf{x}, 2\epsilon)} |f_{\mathbf{X}}(\mathbf{y}) - g(\mathbf{y})| d\mathbf{y} \\ &\leq \frac{2^d}{\eta} (f_{\mathbf{X}} - g)^*(\mathbf{x}) \end{aligned} \quad (100)$$

where $(f_{\mathbf{x}} - g)^*$ denotes the Hardy-Littlewood maximal function for $f_{\mathbf{x}} - g$, i.e.,

$$(f_{\mathbf{x}} - g)^*(\mathbf{x}) \triangleq \sup_{\rho > 0} \frac{1}{\lambda(\mathcal{B}(\mathbf{x}, \rho))} \int_{\mathcal{B}(\mathbf{x}, \rho)} |f_{\mathbf{x}}(\mathbf{y}) - g(\mathbf{y})| d\mathbf{y}, \quad \mathbf{x} \in \mathbb{R}^d. \quad (101)$$

In (100), we have used that, for every $\mathbf{x} \in \mathcal{B}_{i,\epsilon}$ and $i \in \mathcal{I}$, we have $\mathcal{B}_{i,\epsilon} \subseteq \mathcal{B}(\hat{\mathbf{x}}_i, \epsilon) \subseteq \mathcal{B}(\mathbf{x}, 2\epsilon)$ and

$$\Lambda_{i,\epsilon} \geq \eta \lambda(\mathcal{B}(\hat{\mathbf{x}}_i, \epsilon)) = 2^{-d} \lambda(\mathcal{B}(\mathbf{x}, 2\epsilon)).$$

Combining (99) and (100) with (97), we obtain

$$\begin{aligned} & \left| f_{\mathbf{x}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) - f_{\mathbf{x}}(\mathbf{x}) \right| \\ & \leq \sum_{i \in \mathcal{I}} \left| \frac{1}{\Lambda_{i,\epsilon}} \int_{\mathcal{B}_{i,\epsilon}} g(\mathbf{y}) d\mathbf{y} - g(\mathbf{x}) \right| \mathbb{1}\{\mathbf{x} \in \mathcal{B}_{i,\epsilon}\} \\ & \quad + \sum_{i \in \mathcal{I}} \frac{2^d}{\eta} (f_{\mathbf{x}} - g)^*(\mathbf{x}) \mathbb{1}\{\mathbf{x} \in \mathcal{B}_{i,\epsilon}\} + \sum_{i \in \mathcal{I}} |f_{\mathbf{x}}(\mathbf{x}) - g(\mathbf{x})| \mathbb{1}\{\mathbf{x} \in \mathcal{B}_{i,\epsilon}\} \\ & \leq \sum_{i \in \mathcal{I}} \left| \frac{1}{\Lambda_{i,\epsilon}} \int_{\mathcal{B}_{i,\epsilon}} g(\mathbf{y}) d\mathbf{y} - g(\mathbf{x}) \right| \mathbb{1}\{\mathbf{x} \in \mathcal{B}_{i,\epsilon}\} + \frac{2^d}{\eta} (f_{\mathbf{x}} - g)^*(\mathbf{x}) + |f_{\mathbf{x}}(\mathbf{x}) - g(\mathbf{x})|, \quad \mathbf{x} \in \mathbb{R}^d \end{aligned} \quad (102)$$

since the sets $\mathcal{B}_{i,\epsilon}$, $i \in \mathcal{I}$ are disjoint. The second and third term on the RHS of (102) are independent of D and $q(\cdot)$. The first term on the RHS of (102) vanishes as D tends to zero for any sequence of quantizers. Indeed, the continuity of g implies that for every $\vartheta > 0$ and $\mathbf{x} \in \mathbb{R}^d$ there exists an $\epsilon_0 > 0$ such that

$$|g(\mathbf{y}) - g(\mathbf{x})| \leq \vartheta, \quad \text{for } \|\mathbf{x} - \mathbf{y}\| \leq 2\epsilon_0. \quad (103)$$

Since $\mathbf{x}, \mathbf{y} \in \mathcal{B}_{i,\epsilon}$ satisfy $\|\mathbf{x} - \mathbf{y}\| \leq 2\epsilon$, it follows that for every $\vartheta > 0$ and $\mathbf{x} \in \mathbb{R}^d$ there exists an $\epsilon_0 > 0$ such that

$$\left| \frac{1}{\Lambda_{i,\epsilon}} \int_{\mathcal{B}_{i,\epsilon}} g(\mathbf{y}) d\mathbf{y} - g(\mathbf{x}) \right| \mathbb{1}\{\mathbf{x} \in \mathcal{B}_{i,\epsilon}\} \leq \vartheta \mathbb{1}\{\mathbf{x} \in \mathcal{B}_{i,\epsilon}\}, \quad \epsilon \leq \epsilon_0. \quad (104)$$

Using that the sets $\mathcal{B}_{i,\epsilon}$, $i \in \mathcal{I}$ are disjoint, we conclude that for every $\vartheta > 0$ and $\mathbf{x} \in \mathbb{R}^d$ there exists an $\epsilon_0 > 0$ such that

$$\sum_{i \in \mathcal{I}} \left| \frac{1}{\Lambda_{i,\epsilon}} \int_{\mathcal{B}_{i,\epsilon}} g(\mathbf{y}) d\mathbf{y} - g(\mathbf{x}) \right| \mathbb{1}\{\mathbf{x} \in \mathcal{B}_{i,\epsilon}\} \leq \vartheta, \quad \epsilon \leq \epsilon_0. \quad (105)$$

Since $\vartheta > 0$ is arbitrary and ϵ vanishes as $D \rightarrow 0$, this implies that for every $\mathbf{x} \in \mathbb{R}^d$ and any sequence of quantizers

$$\lim_{D \downarrow 0} \sum_{i \in \mathcal{I}} \left| \frac{1}{\Lambda_{i,\epsilon}} \int_{\mathcal{B}_{i,\epsilon}} g(\mathbf{y}) d\mathbf{y} - g(\mathbf{x}) \right| \mathbb{1}\{\mathbf{x} \in \mathcal{B}_{i,\epsilon}\} = 0. \quad (106)$$

We conclude the proof of Lemma 6 by applying (102) and (106) to upper-bound the Lebesgue measure on the LHS of (95). Indeed, we have

$$\begin{aligned} & \lambda \left(\left\{ \mathbf{x} \in \mathbb{R}^d : \overline{\lim}_{D \downarrow 0} \left| f_{\mathbf{x}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) - f_{\mathbf{x}}(\mathbf{x}) \right| > 2\xi \right\} \right) \\ & \leq \lambda \left(\left\{ \mathbf{x} \in \mathbb{R}^d : \frac{2^d}{\eta} (f_{\mathbf{x}} - g)^*(\mathbf{x}) + |f_{\mathbf{x}}(\mathbf{x}) - g(\mathbf{x})| > 2\xi \right\} \right) \\ & \leq \lambda \left(\left\{ \mathbf{x} \in \mathbb{R}^d : \frac{2^d}{\eta} (f_{\mathbf{x}} - g)^*(\mathbf{x}) > \xi \right\} \right) + \lambda \left(\left\{ \mathbf{x} \in \mathbb{R}^d : |f_{\mathbf{x}}(\mathbf{x}) - g(\mathbf{x})| > \xi \right\} \right). \end{aligned} \quad (107)$$

The first term on the RHS of (107) can be upper-bounded by using the Hardy-Littlewood maximal inequality [24, Theorem 3.4, p. 55]

$$\lambda \left(\left\{ \mathbf{x} \in \mathbb{R}^d : \frac{2^d}{\eta} (f_{\mathbf{x}} - g)^*(\mathbf{x}) > \xi \right\} \right) \leq \frac{2^d \alpha_d}{\eta \xi} \int |f_{\mathbf{x}}(\mathbf{x}) - g(\mathbf{x})| d\mathbf{x} \quad (108)$$

for some constant α_d that only depends on d . Likewise, the second term on the RHS of (107) can be upper-bounded using Chebyshev's inequality [23, Theorem 4.10.7, p. 192]

$$\lambda \left(\left\{ \mathbf{x} \in \mathbb{R}^d : |f_{\mathbf{x}}(\mathbf{x}) - g(\mathbf{x})| > \xi \right\} \right) \leq \frac{1}{\xi} \int |f_{\mathbf{x}}(\mathbf{x}) - g(\mathbf{x})| d\mathbf{x}. \quad (109)$$

Combining (108) and (109) with (98) and (107), it follows that

$$\lambda \left(\left\{ \mathbf{x} \in \mathbb{R}^d : \overline{\lim}_{D \downarrow 0} |f_{\mathbf{x}}^{(\Lambda)}(\mathbf{x}; \{\mathcal{B}_{i,\epsilon}\}) - f_{\mathbf{x}}(\mathbf{x})| > 2\xi \right\} \right) \leq \frac{1 + 2^d \alpha_d / \eta}{\xi} \epsilon. \quad (110)$$

This proves (95) upon letting ϵ tend to zero from above, which was the last step required to prove Lemma 6.

APPENDIX B

PROOF OF THEOREM 7

Following the steps (51)–(61) in the proof of Theorem 1 in Section IV-B particularized for $d = 1$, we obtain that

$$\begin{aligned} H(q(X)) - R(D) &\geq \frac{1}{r} \log(r 2^r \Gamma(1 + 1/r)^r e) - \frac{1}{r} \sum_i \Pr(X \in \mathcal{B}_{i,\epsilon}) \log \left(\frac{\Lambda_{i,\epsilon}}{D^{1/r}} + 2\Gamma \left(\frac{1}{r}, \frac{1}{\kappa \delta} \right) \right)^r \\ &\quad - \kappa \log \left(\frac{2}{\kappa^{1/r}} + 2\Gamma(1/r) \right) - \kappa \left| \log \frac{r}{\delta^{1/r}} \right| - \frac{1}{\delta}. \end{aligned} \quad (111)$$

Recall that $\epsilon^r = D/\kappa$. The last three terms on the RHS of (111) are independent of D and vanish as we first let $\kappa \rightarrow 0$ and then $\delta \rightarrow \infty$. To achieve

$$R_{r,1} = \frac{1}{r} \log \left(\frac{\Gamma(1 + 1/r)^r e}{1 + 1/r} \right)$$

a sequence of quantizers (parametrized by D) must therefore satisfy

$$\lim_{\kappa \downarrow 0} \overline{\lim}_{D \downarrow 0} \frac{1}{r} \sum_i \Pr(X \in \mathcal{B}_{i,\epsilon}) \log \left(\frac{\Lambda_{i,\epsilon}}{D^{1/r}} + 2\Gamma \left(\frac{1}{r}, \frac{1}{\kappa \delta} \right) \right)^r \geq \frac{1}{r} \log(2^r(1+r)). \quad (112)$$

(As $\kappa \rightarrow 0$, the term on the LHS of (112) becomes independent of $\delta > 0$.) For the sake of compactness, we shall use in the rest of the proof the following notation:³

Let $V \triangleq 2^r(1+r)$. Further let $v \triangleq 2\Gamma \left(\frac{1}{r}, \frac{1}{\kappa \delta} \right)$, and recall that $\lim_{\kappa \rightarrow 0} v = 0$ for every $\delta > 0$. Define

$$\underline{\mathcal{I}}_D \triangleq \left\{ i \in \mathbb{Z} : \frac{\Lambda_{i,\epsilon}^r}{D} \leq V - v \right\} \quad (113a)$$

$$\overline{\mathcal{I}}_D \triangleq \left\{ i \in \mathbb{Z} : \frac{\Lambda_{i,\epsilon}^r}{D} \geq V + v \right\} \quad (113b)$$

³While all introduced quantities depend on κ , to keep the notation compact we only make the dependence on D explicit.

and

$$\underline{q}_D \triangleq \frac{1}{1 - \wp_\epsilon} \sum_{i \in \underline{\mathcal{I}}_D} \Pr(X \in \mathcal{B}_{i,\epsilon}), \quad \underline{\mu}_D \triangleq \frac{1}{(1 - \wp_\epsilon)\underline{q}_D} \sum_{i \in \underline{\mathcal{I}}_D} \Pr(X \in \mathcal{B}_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^r}{D} \quad (114a)$$

$$\bar{q}_D \triangleq \frac{1}{1 - \wp_\epsilon} \sum_{i \in \bar{\mathcal{I}}_D} \Pr(X \in \mathcal{B}_{i,\epsilon}), \quad \bar{\mu}_D \triangleq \frac{1}{(1 - \wp_\epsilon)\bar{q}_D} \sum_{i \in \bar{\mathcal{I}}_D} \Pr(X \in \mathcal{B}_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^r}{D} \quad (114b)$$

$$\underline{\bar{q}}_D \triangleq \frac{1}{1 - \wp_\epsilon} \sum_{i \in \mathbb{Z} \setminus (\underline{\mathcal{I}}_D \cup \bar{\mathcal{I}}_D)} \Pr(X \in \mathcal{B}_{i,\epsilon}), \quad \underline{\bar{\mu}}_D \triangleq \frac{1}{(1 - \wp_\epsilon)\underline{\bar{q}}_D} \sum_{i \in \mathbb{Z} \setminus (\underline{\mathcal{I}}_D \cup \bar{\mathcal{I}}_D)} \Pr(X \in \mathcal{B}_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^r}{D} \quad (114c)$$

where \wp_ϵ was defined in (58). Finally, define

$$\mu_D \triangleq \underline{q}_D \underline{\mu}_D + \bar{q}_D \bar{\mu}_D + \underline{\bar{q}}_D \underline{\bar{\mu}}_D. \quad (115)$$

By definition of $\underline{\mathcal{I}}$ and $\bar{\mathcal{I}}$, we have

$$\underline{\mu}_D \leq V - \vartheta \quad \text{and} \quad \bar{\mu}_D \geq V + \vartheta. \quad (116)$$

Furthermore, by Lemma 6 and (59),

$$\lim_{\kappa \downarrow 0} \overline{\lim_{D \downarrow 0}} \mu_D \leq V. \quad (117)$$

Consequently, for any arbitrary $\varepsilon > 0$, there exist κ_0 and D_0 such that

$$\mu_D \leq V + \varepsilon, \quad (\kappa \leq \kappa_0, D \leq D_0). \quad (118)$$

Without loss of generality, we implicitly assume that κ and D are sufficiently small, so that (118) holds.

We next apply steps similar to (62) and (63) to upper-bound

$$\begin{aligned} & \frac{1}{1 - \wp_\epsilon} \sum_i \Pr(X \in \mathcal{B}_{i,\epsilon}) \log \left(\frac{\Lambda_{i,\epsilon}}{D^{1/r}} + v \right)^r \\ & \leq \underline{q}_D \log \left(\underline{\mu}_D + v^r \right) + \bar{q}_D \log \left(\bar{\mu}_D + v^r \right) + \underline{\bar{q}}_D \log \left(\underline{\bar{\mu}}_D + v^r \right), \quad \text{for } r < 1 \end{aligned} \quad (119a)$$

and

$$\begin{aligned} & \frac{1}{1 - \wp_\epsilon} \sum_i \Pr(X \in \mathcal{B}_{i,\epsilon}) \log \left(\frac{\Lambda_{i,\epsilon}}{D^{1/r}} + v \right)^r \\ & \leq r \left[\underline{q}_D \log \left(\underline{\mu}_D^{1/r} + v \right) + \bar{q}_D \log \left(\bar{\mu}_D^{1/r} + v \right) + \underline{\bar{q}}_D \log \left(\underline{\bar{\mu}}_D^{1/r} + v \right) \right], \quad \text{for } r \geq 1. \end{aligned} \quad (119b)$$

It follows that, for $r < 1$, any sequence of quantizers satisfying (112) must also satisfy

$$\lim_{\kappa \downarrow 0} \overline{\lim_{D \downarrow 0}} \left\{ \underline{q}_D \log \left(\underline{\mu}_D + v^r \right) + \bar{q}_D \log \left(\bar{\mu}_D + v^r \right) + \underline{\bar{q}}_D \log \left(\underline{\bar{\mu}}_D + v^r \right) - \log(V + v^r) \right\} \geq 0. \quad (120a)$$

Likewise, for $r \geq 1$, any sequence of quantizers satisfying (112) must also satisfy

$$\lim_{\kappa \downarrow 0} \overline{\lim_{D \downarrow 0}} \left\{ \underline{q}_D \log \left(\underline{\mu}_D^{1/r} + v \right) + \bar{q}_D \log \left(\bar{\mu}_D^{1/r} + v \right) + \underline{\bar{q}}_D \log \left(\underline{\bar{\mu}}_D^{1/r} + v \right) - \log(V^{1/r} + v) \right\} \geq 0. \quad (120b)$$

We conclude the proof of Theorem 7 for the case $r \geq 1$ by demonstrating that any sequence of quantizers satisfying (120b) must satisfy

$$\lim_{\kappa \downarrow 0} \overline{\lim_{D \downarrow 0}} \bar{q}_D = 1, \quad \text{for every } \vartheta > 0. \quad (121)$$

Substituting $\rho = 1/\kappa$, this can be written as

$$\lim_{\rho \rightarrow \infty} \overline{\lim}_{D \downarrow 0} \frac{1}{1 - \wp_\epsilon} \sum_i \Pr(X \in \mathcal{B}_{i, \rho D^{1/r}}) \mathbb{1} \left\{ \left| \frac{\Lambda_{i, \rho D^{1/r}}^r}{D} - 2^r(1+r) \right| \leq \vartheta \right\} = 1, \quad \text{for every } \vartheta > 0 \quad (122)$$

which by Lemma 5 is equivalent to (69). The proof for $r < 1$ is almost identical and is therefore omitted.

To prove (121) we use that, by the strict concavity of $x \mapsto \log x$, there exists a linear function $x \mapsto \ell_{x_0}(x)$ such that

$$\log(x+v) \leq \ell_{x_0}(x), \quad x \geq 0 \quad (123)$$

with equality if, and only if, $x = x_0$. (Specifically, $\ell_{x_0}(x) = \frac{x+v}{x_0+v} \log(x_0+v) - 1$.) Moreover, we have

$$\log(\mu_D^{1/r} + v) \geq \underline{q}_D \ell_{\mu_D^{1/r}}(\mu_D^{1/r}) + \bar{q}_D \ell_{\mu_D^{1/r}}(\bar{\mu}_D^{1/r}) + \bar{q}_D \ell_{\mu_D^{1/r}}(\bar{\mu}_D^{1/r}) \quad (124)$$

since $x_0 \mapsto \log(x_0+v) - \mathbb{E}[\ell_{x_0}(X)]$ (for any discrete random variable X) is monotonically increasing in x_0 and nonnegative for $x_0 \geq \mathbb{E}[X]$, and since $\underline{q}_D \mu_D^{1/r} + \bar{q}_D \bar{\mu}_D^{1/r} + \bar{q}_D \bar{\mu}_D^{1/r} \leq \mu_D^{1/r}$. The LHS of (120b) can thus be upper-bounded by

$$\begin{aligned} \underline{q}_D \left[\log(\mu_D^{1/r} + v) - \ell_{\mu_D^{1/r}}(\mu_D^{1/r}) - \log\left(\frac{V^{1/r} + v}{\mu_D^{1/r} + v}\right) \right] \\ + \bar{q}_D \left[\log(\bar{\mu}_D^{1/r} + v) - \ell_{\mu_D^{1/r}}(\bar{\mu}_D^{1/r}) - \log\left(\frac{V^{1/r} + v}{\mu_D^{1/r} + v}\right) \right] \\ + \bar{q}_D \left[\log(\bar{\mu}_D^{1/r} + v) - \ell_{\mu_D^{1/r}}(\bar{\mu}_D^{1/r}) - \log\left(\frac{V^{1/r} + v}{\mu_D^{1/r} + v}\right) \right]. \end{aligned} \quad (125)$$

By (117) and (123), the third term in (125) satisfies

$$\overline{\lim}_{\kappa \downarrow 0} \overline{\lim}_{D \downarrow 0} \bar{q}_D \left[\log(\bar{\mu}_D^{1/r} + v) - \ell_{\mu_D^{1/r}}(\bar{\mu}_D^{1/r}) - \log\left(\frac{V^{1/r} + v}{\mu_D^{1/r} + v}\right) \right] \leq 0. \quad (126)$$

We further have

$$\begin{aligned} \log(\mu_D^{1/r} + v) - \ell_{\mu_D^{1/r}}(\mu_D^{1/r}) - \log\left(\frac{V^{1/r} + v}{\mu_D^{1/r} + v}\right) \\ \leq \log((V - \vartheta)^{1/r} + v) - \ell_{(V+\varepsilon)^{1/r}}((V - \vartheta)^{1/r}) + \log\left(\frac{(V + \varepsilon)^{1/r} + v}{V^{1/r} + v}\right) \\ \triangleq \underline{K} \end{aligned} \quad (127)$$

and

$$\begin{aligned} \log(\bar{\mu}_D^{1/r} + v) - \ell_{\mu_D^{1/r}}(\bar{\mu}_D^{1/r}) - \log\left(\frac{V^{1/r} + v}{\mu_D^{1/r} + v}\right) \\ \leq \log((V + \vartheta)^{1/r} + v) - \ell_{(V+\varepsilon)^{1/r}}((V + \vartheta)^{1/r}) + \log\left(\frac{(V + \varepsilon)^{1/r} + v}{V^{1/r} + v}\right) \\ \triangleq \bar{K}. \end{aligned} \quad (128)$$

Here, we used (116) and (118) together with the facts that $x \mapsto \ell_{x_0}(x) - \log(x+v)$ is monotonically decreasing for $x \leq x_0$ and monotonically increasing for $x \geq x_0$, and $x_0 \mapsto \log(x_0+v) - \ell_{x_0}(x)$ is monotonically increasing.

Combining (125)–(128), it follows that (120b) can only be satisfied if

$$\lim_{\kappa \downarrow 0} \lim_{D \downarrow 0} \max \{ \underline{K}, \overline{K} \} \left(\underline{q}_D + \overline{q}_D \right) \geq 0. \quad (129)$$

Since for $\varepsilon > 0$ sufficiently small, we have

$$\lim_{\kappa \downarrow 0} \max \{ \underline{K}, \overline{K} \} < 0 \quad (130)$$

the condition (129), in turn, can only be satisfied if

$$\lim_{\kappa \downarrow 0} \lim_{D \downarrow 0} \left(\underline{q}_D + \overline{q}_D \right) = 0. \quad (131)$$

Using that $\overline{q}_D = 1 - \underline{q}_D - \overline{q}_D$, the claim (121) follows. This concludes the proof of Theorem 7.

ACKNOWLEDGMENT

Stimulating discussions with Tamás Linder and Ram Zamir are gratefully acknowledged. The authors further wish to thank Giuseppe Durisi for calling their attention to reference [12].

REFERENCES

- [1] C. E. Shannon, “Coding theorems for a discrete source with a fidelity criterion,” *IRE International Convention Record*, vol. 7, pp. 142–163, 1959.
- [2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. John Wiley & Sons, 1991.
- [3] N. Alon and A. Orlitsky, “A lower bound on the expected length of one-to-one codes,” *IEEE Trans. Inf. Theory*, vol. 40, no. 5, pp. 1670–1672, Sept. 1994.
- [4] H. Gish and J. N. Pierce, “Asymptotically efficient quantizing,” *IEEE Trans. Inf. Theory*, vol. 14, no. 5, pp. 676–683, Sept. 1968.
- [5] R. M. Gray and D. L. Neuhoff, “Quantization,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.
- [6] P. L. Zador, “Topics in the asymptotic quantization of continuous random variables,” Bell Laboratories, Tech. Rep., 1966.
- [7] Y. N. Linkov, “Evaluation of epsilon entropy of random variables for small epsilon,” *Problemy Peredachi Informatsii (Problems of Inform. Transm.)*, vol. 1, pp. 12–18, 1965.
- [8] T. Linder and R. Zamir, “On the asymptotic tightness of the Shannon lower bound,” *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 2026–2031, Nov. 1994.
- [9] T. Koch, “The Shannon lower bound is asymptotically tight,” *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6155–6161, Nov. 2016.
- [10] R. M. Gray, T. Linder, and J. Li, “A Lagrangian formulation of Zador’s entropy-constrained quantization theorem,” *IEEE Trans. Inf. Theory*, vol. 28, no. 3, pp. 695–707, Mar. 2002.
- [11] A. György and T. Linder, “On the structure of optimal entropy-constrained scalar quantizers,” *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 416–427, Feb. 2002.
- [12] Y. Wu and S. Verdú, “Rényi information dimension: Fundamental limits of almost lossless analog compression,” *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3721–3748, Aug. 2010.
- [13] A. Rényi, “On the dimension and entropy of probability distributions,” *Acta Mathematica Hungarica*, vol. 10, no. 1–2, Mar. 1959.
- [14] T. Kawabata and A. Dembo, “The rate-distortion dimension of sets and measures,” *IEEE Trans. Inf. Theory*, vol. 40, no. 5, pp. 1564–1572, Sept. 1994.
- [15] T. Linder and K. Zeger, “Asymptotic entropy-constrained performance of tessellating and universal randomized lattice quantization,” *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 575–579, Mar. 1994.
- [16] Y. Yamada, S. Tazaki, and R. M. Gray, “Asymptotic performance of block quantizers with difference distortion measures,” *IEEE Trans. Inf. Theory*, vol. 26, pp. 6–14, Jan. 1980.
- [17] V. Kostina, “Data compression with low distortion and finite blocklength,” in *Proc. 53rd Allerton Conf. Comm., Contr. and Comp.*, Allerton H., Monticello, IL, Sep. 30 – Oct. 2, 2015.
- [18] —, “Data compression with low distortion and finite blocklength,” *IEEE Trans. Inf. Theory*, Mar. 2017, to be published.

- [19] P. Billingsley, *Probability and Measure*, 3rd ed., ser. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, 1995.
- [20] R. B. Ash, *Information Theory*, ser. Dover Books on Mathematics. Mineola, NY: Courier Dover Publications, 1990.
- [21] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.
- [22] J. H. Conway and N. J. A. Sloane, "A lower bound on the average error of vector quantizers," *IEEE Trans. Inf. Theory*, vol. 31, no. 1, pp. 106–109, Jan. 1985.
- [23] R. B. Ash and C. A. Doléans-Dade, *Probability and Measure Theory*, 2nd ed. Elsevier/Academic Press, 2000.
- [24] E. M. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton University Press, 1971.